

Deep Learning for Medical Applications

Nataliya Sokolovska

Sorbonne University
Paris, France

Master 2 in Statistics
March, 22, 2019

Outline

Motivation: Deep Learning

Some Theoretical Foundations (very brief)

Some Applications in Biology/Medicine

Automatic Immune Cells Detection

Predicting specificities of DNA- and RNA-binding proteins

Pulmonary Nodules and Computer-Aided Diagnosis

Deep Learning Results on the MicrObese Data

Results of a Deep Kernel Approach

Deep Self-Organising Maps

Perspectives

Motivation: Deep Learning

Some Theoretical Foundations (very brief)

Some Applications in Biology/Medicine

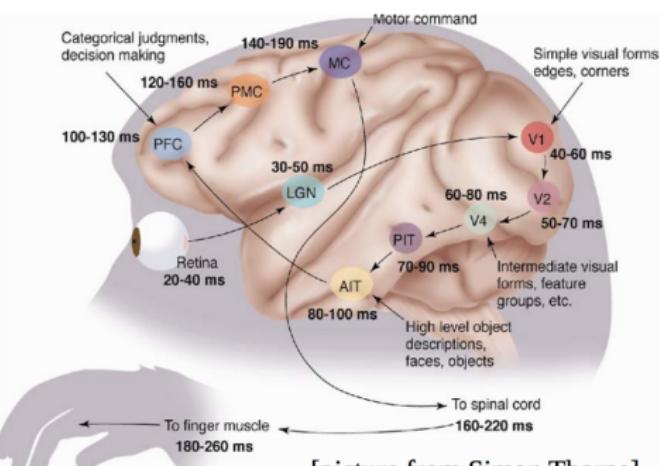
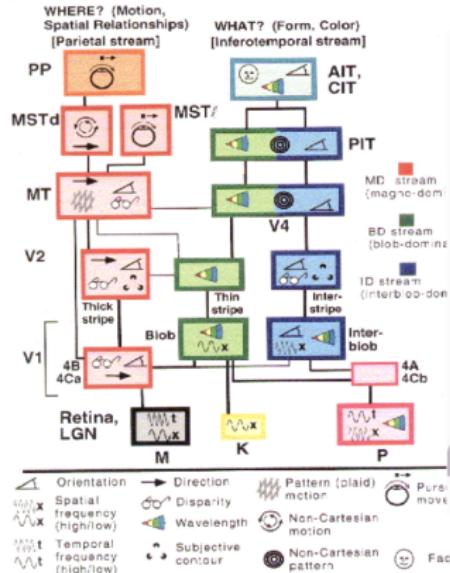
Deep Learning Results on the MicrObese Data

Perspectives

“Efficient Computer”

- ▶ $\approx 10^{11}$ neurons
- ▶ 10^{16} “operations” per second

- The ventral (recognition) pathway in the visual cortex has multiple stages
- Retina - LGN - V1 - V2 - V4 - PIT - AIT



[picture from Simon Thorpe]

[Gallant & Van Essen]

from Y. LeCun

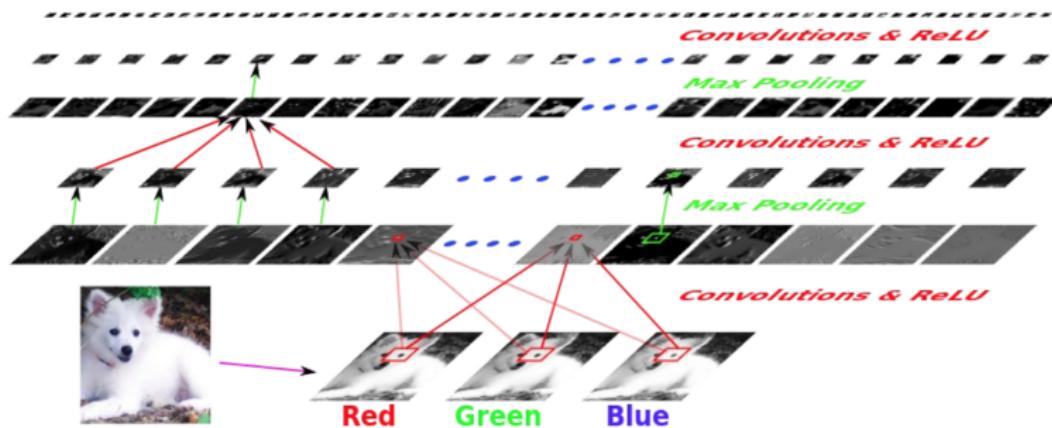
Deep Learning

- ▶ Deep Learning is not a brand new idea
- ▶ Artificial Neural Networks (a biologically inspired model)
 - ▶ Non-linear separator (often more efficient than a linear one)
 - ▶ Supervised learning, clustering, dimensionality reduction, structured prediction, ...
- ▶ Deep learning = learning hierarchical representations
- ▶ Deep learning is about modelling high-level abstractions in data by using multiple processing layers
- ▶ Deep learning is used for various applications (natural language processing, bioinformatics), especially powerful for image processing

An Example of Deep Architecture for Image Processing

■ 1 to 10 billion connections, 10 million to 1 billion parameters, 8 to 20 layers.

Samoyed (1.6); Papillon (5.7); Pomeranian (2.7); Arctic Fox (1.0); Eskimo Dog (0.6); White Wolf (0.4); Siberian Husky (0.4)



from Y. LeCun, Y. Bengio, and G. Hinton. Deep Learning, Nature, 2015

Motivation: Deep Learning

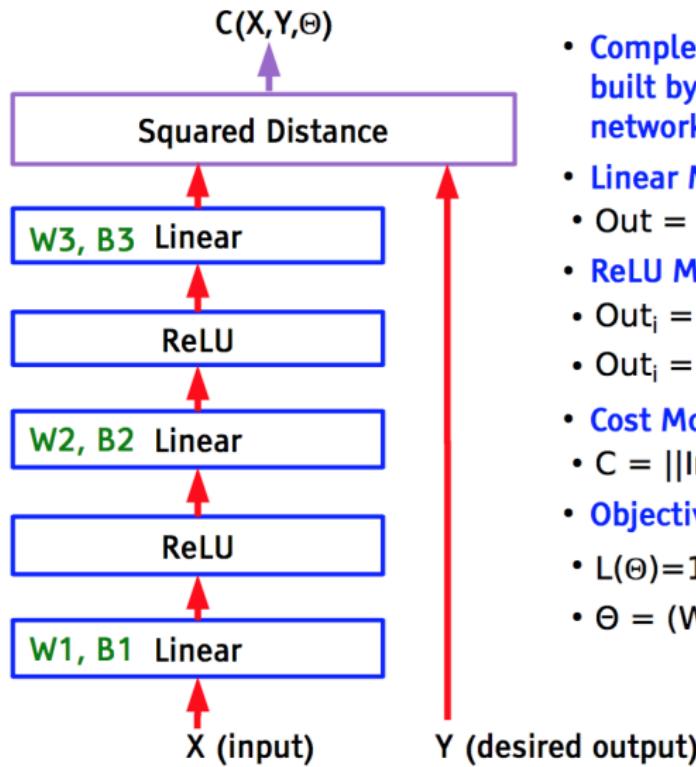
Some Theoretical Foundations (very brief)

Some Applications in Biology/Medicine

Deep Learning Results on the MicrObese Data

Perspectives

Typical Multilayer Neural Network Architecture

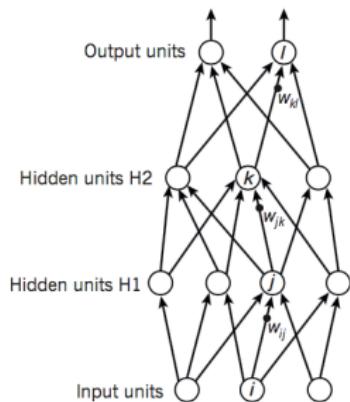


- Complex learning machines can be built by assembling modules into networks
- Linear Module
 - $\text{Out} = W \cdot \text{In} + B$
- ReLU Module (Rectified Linear Unit)
 - $\text{Out}_i = 0$ if $\text{In}_i < 0$
 - $\text{Out}_i = \text{In}_i$ otherwise
- Cost Module: Squared Distance
 - $C = ||\text{In}_1 - \text{In}_2||^2$
- Objective Function
 - $L(\Theta) = 1/p \sum_k C(X^k, Y^k, \Theta)$
- $\Theta = (W1, B1, W2, B2, W3, B3)$

from Y. LeCun

Inference in Deep Networks: Forward and Backward Passes

c



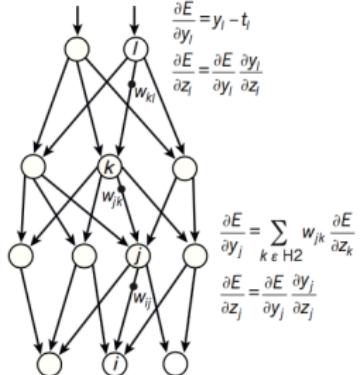
$$y_l = f(z_l)$$
$$z_l = \sum_{k \in H2} w_{kl} y_k$$

$$y_k = f(z_k)$$
$$z_k = \sum_{j \in H1} w_{jk} y_j$$

$$y_j = f(z_j)$$
$$z_j = \sum_{i \in \text{Input}} w_{ij} x_i$$

d

Compare outputs with correct answer to get error derivatives



from Y. LeCun, Y. Bengio, and G. Hinton. Deep Learning, Nature, 2015

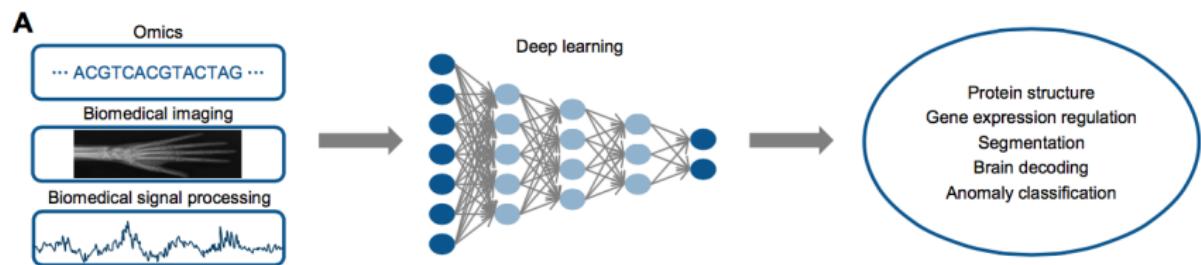
Some Applications in Biology/Medicine

Automatic Immune Cells Detection

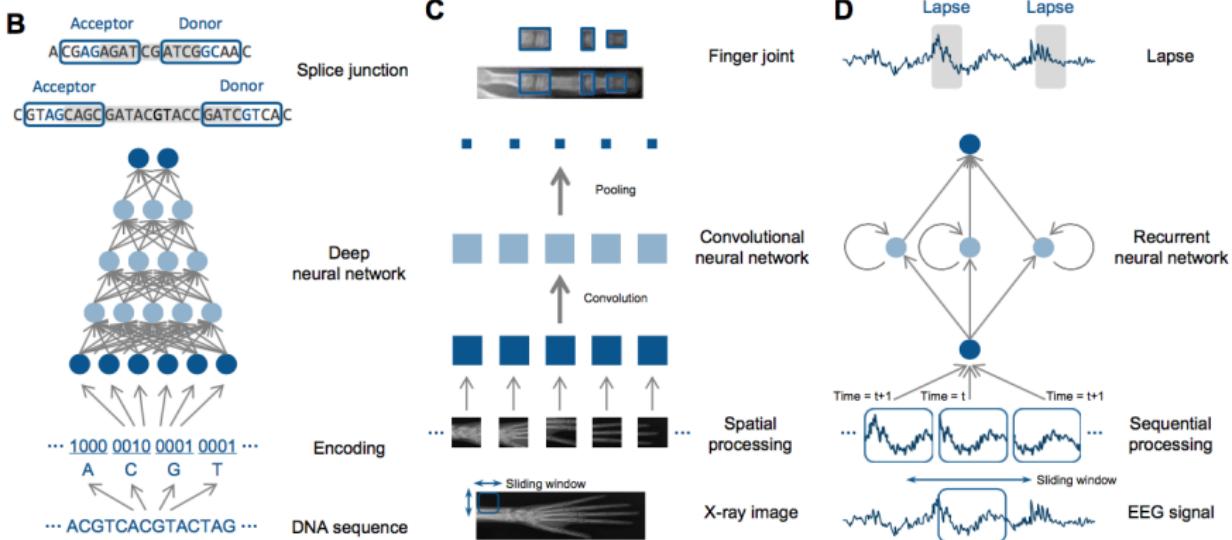
Predicting specificities of DNA- and RNA-binding proteins

Pulmonary Nodules and Computer-Aided Diagnosis

Application of Deep Learning in Bioinformatics Research



from S. Min, B. Lee, and S. Yoon. Deep Learning in Bioinformatics.
<https://arxiv.org/pdf/1603.06430.pdf>



- ▶ **B** Omics domain. Prediction of splice junctions in DNA sequence data with deep neural network
- ▶ **C** Biomedical imaging. Finger joint detection from X-ray images with convolutional neural network
- ▶ **D** Biomedical signal processing. Lapse detection from EEG signal with recurrent neural network

Motivation: Deep Learning

Some Theoretical Foundations (very brief)

Some Applications in Biology/Medicine

Automatic Immune Cells Detection

Predicting specificities of DNA- and RNA-binding proteins
Pulmonary Nodules and Computer-Aided Diagnosis

Deep Learning Results on the MicrObese Data

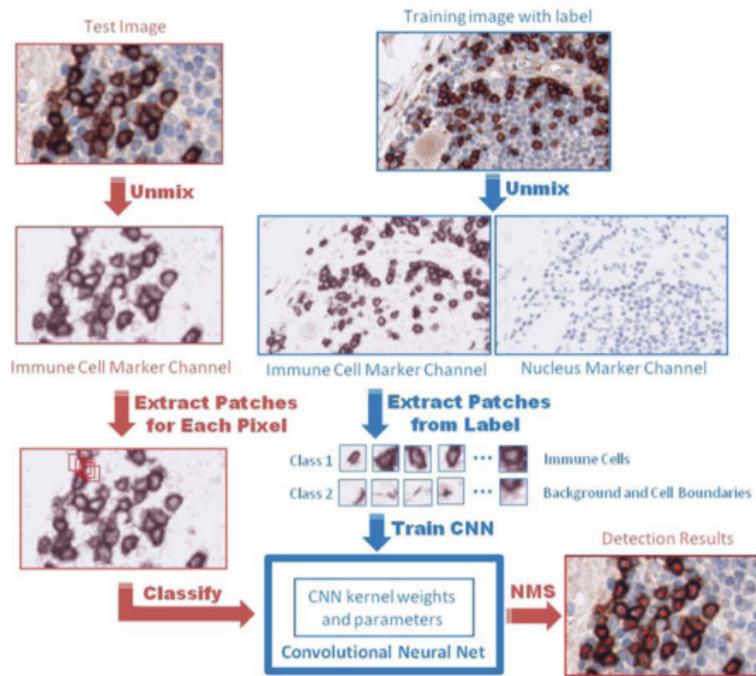
Perspectives

Automatic Immune Cells Detection

- ▶ Immunohistochemistry (IHC) can be used to determine the distribution and localisation of the differentially expressed biomarkers of immune cells (T-cells, B-cells)
- ▶ Manually count each subset of immune cells is tedious and time consuming; automatic detection is very attractive
- ▶ A novel method for automatic immune cell counting on digitally scanned images of IHC stained slides.

T. Cheng and C. Chefd'hotel. Deep Learning Based Automatic Immune Cell Detection for Immunohistochemistry Images. Workshop on Machine Learning in Medical Imaging. 2014

Automatic Immune Cells Detection



T. Cheng and C. Chefd'hotel. Deep Learning Based Automatic Immune Cell Detection for Immunohistochemistry Images. Workshop on Machine Learning in Medical Imaging. 2014

Motivation: Deep Learning

Some Theoretical Foundations (very brief)

Some Applications in Biology/Medicine

Automatic Immune Cells Detection

Predicting specificities of DNA- and RNA-binding proteins

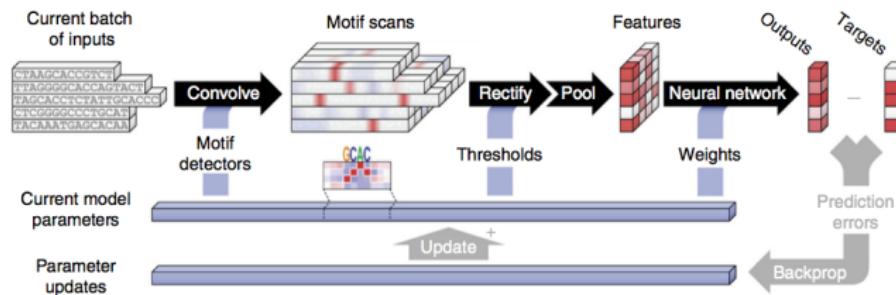
Pulmonary Nodules and Computer-Aided Diagnosis

Deep Learning Results on the MicrObese Data

Perspectives

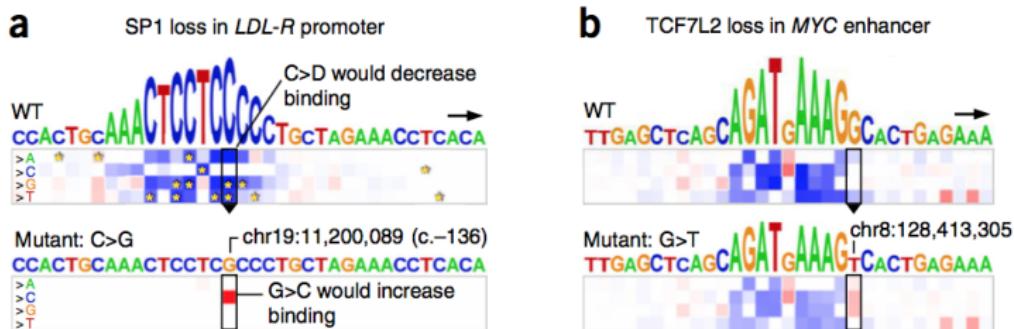
Predicting the Sequence Specificities of DNA- and RNA-binding Proteins by Deep Learning

- ▶ Knowing the sequence specificities of DNA- and RNA-binding proteins is important for developing models of the regulatory processes and for identifying causal disease variants
- ▶ Specificities determined by DeepBind are readily visualized as a weighted ensemble of position weight matrices or as a “mutation map” that indicates how variations affect binding within a specific sequence



B. Alipanahi et al. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*. 2015

Predicting the Sequence Specificities of DNA- and RNA-binding Proteins by Deep Learning



Analysis of potentially disease-causing genomic variants. DeepBind mutation maps were used to understand disease-causing SNVs (single-nucleotide variants) associated with transcription factor binding. (a) A disrupted SP1 binding site in the *LDL-R* promoter that leads to familial hypercholesterolemia. (b) A cancer risk variant in a *MYC* enhancer weakens a TCF7L2 binding site.

Motivation: Deep Learning

Some Theoretical Foundations (very brief)

Some Applications in Biology/Medicine

Automatic Immune Cells Detection

Predicting specificities of DNA- and RNA-binding proteins

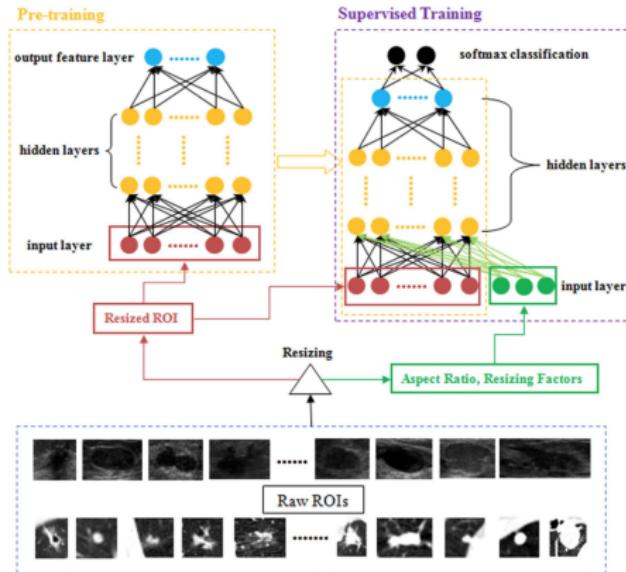
Pulmonary Nodules and Computer-Aided Diagnosis

Deep Learning Results on the MicrObese Data

Perspectives

Computer-Aided Diagnosis with Deep Learning Architecture

- The differential diagnosis of benign and malignant nodules/lesions



Motivation: Deep Learning

Some Theoretical Foundations (very brief)

Some Applications in Biology/Medicine

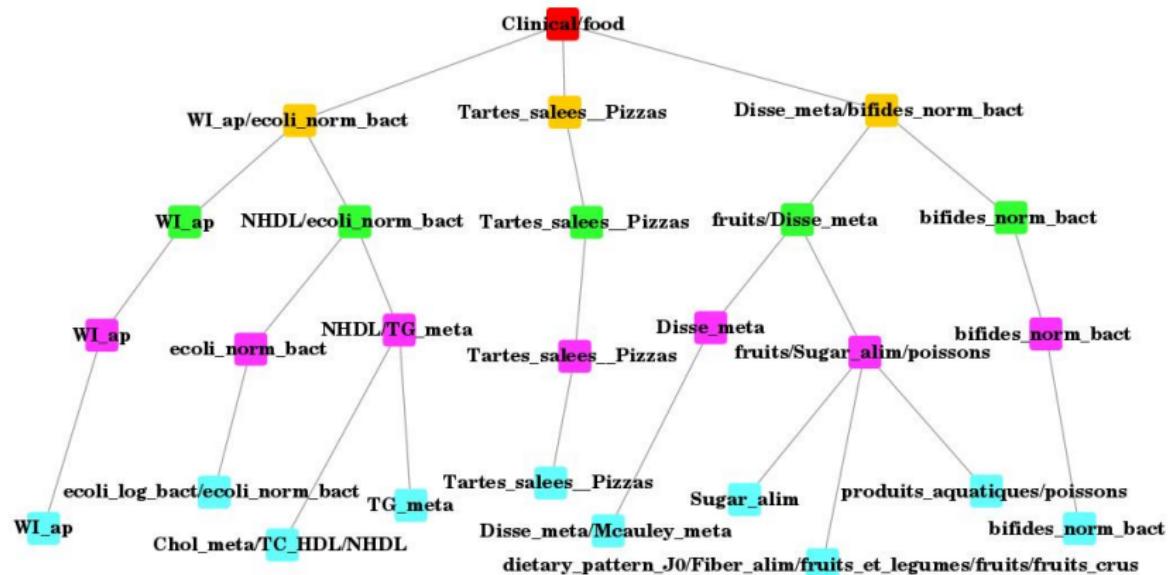
Deep Learning Results on the MicrObese Data

Results of a Deep Kernel Approach

Deep Self-Organising Maps

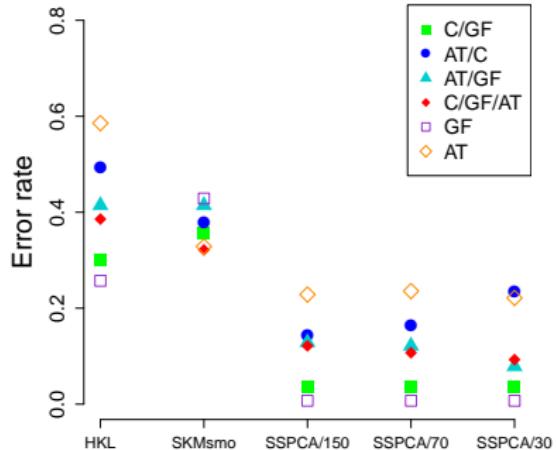
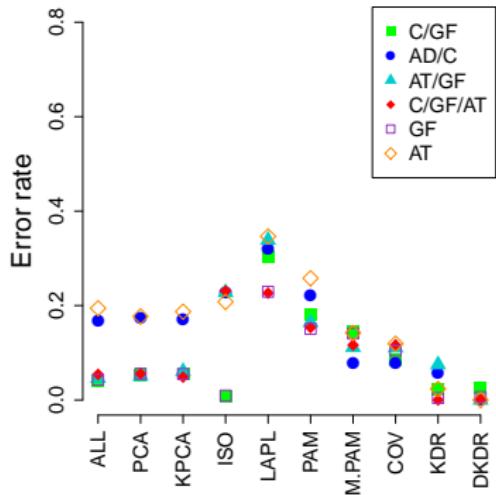
Perspectives

Deep Kernel Dimensionality Reduction



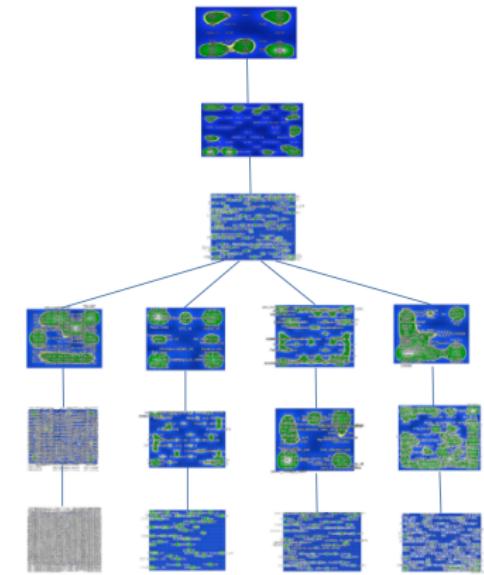
- ▶ N. Sokolovska, K. Clément, J.-D. Zucker. Deep Kernel Dimensionality Reduction for Scalable Data Integration. International Journal of Approximate Reasoning, 2016
- ▶ N. Sokolovska, S. Rizkalla, K. Clément, J.-D. Zucker. Continuous and Discrete Deep Classifiers for Data Integration. International Symposium on Intelligent Data Analysis (IDA), 2015

Deep Kernel Dimensionality Reduction



- ▶ N. Sokolovska, K. Clément, J.-D. Zucker. Deep Kernel Dimensionality Reduction for Scalable Data Integration. *International Journal of Approximate Reasoning*, 2016
- ▶ N. Sokolovska, S. Rizkalla, K. Clément, J.-D. Zucker. Continuous and Discrete Deep Classifiers for Data Integration. *International Symposium on Intelligent Data Analysis (IDA)*, 2015

Deep Self-Organising Maps for Efficient Heterogeneous Biomedical Signatures Extraction

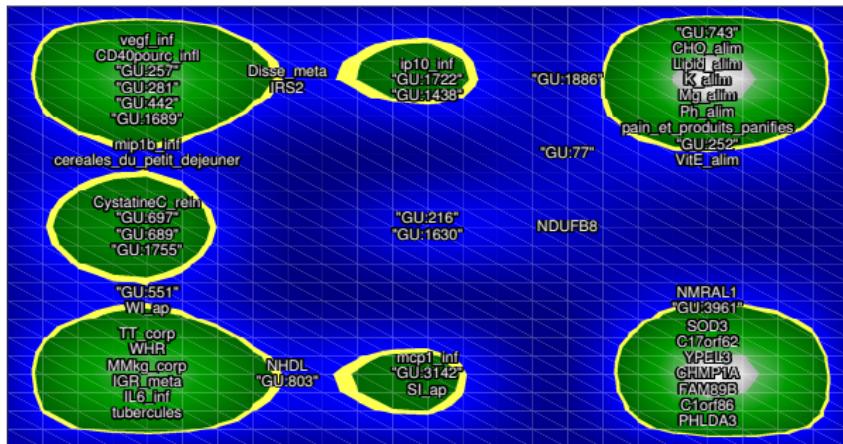


MGS, host, environmental, AT genes (MicrObese Data at T0)

N. Sokolovska, H. T. Nguyen, K. Clément, J.-D. Zucker. Deep Self-Organising Maps for Efficient Heterogeneous Biomedical Signatures Extraction. International Joint Conference on Neural Networks (IJCNN), 2016, accepted.

Deep Self-Organising Maps

A heterogeneous signature that differentiates HGC from LGC



N. Sokolovska, H. T. Nguyen, K. Clément, J.-D. Zucker. Deep Self-Organising Maps for Efficient Heterogeneous Biomedical Signatures Extraction. International Joint Conference on Neural Networks (IJCNN), 2016, accepted.

Deep Self-Organising Maps

A heterogeneous signature of HGC



Deep Self-Organising Maps

A heterogeneous signature of LGC



N. Sokolovska, H. T. Nguyen, K. Clément, J.-D. Zucker. Deep Self-Organising Maps for Efficient Heterogeneous Biomedical Signatures Extraction. International Joint Conference on Neural Networks (IJCNN), 2016, accepted.

Motivation: Deep Learning

Some Theoretical Foundations (very brief)

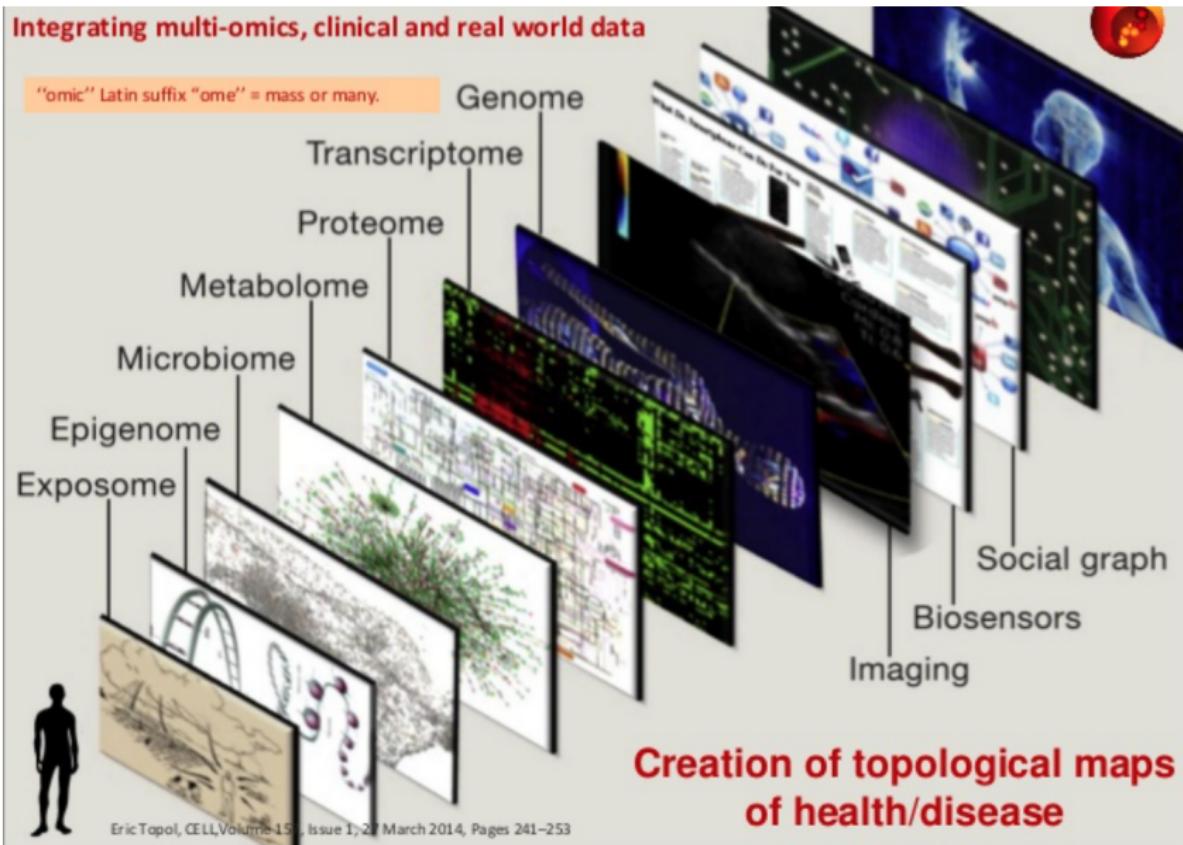
Some Applications in Biology/Medicine

Deep Learning Results on the MicrObese Data

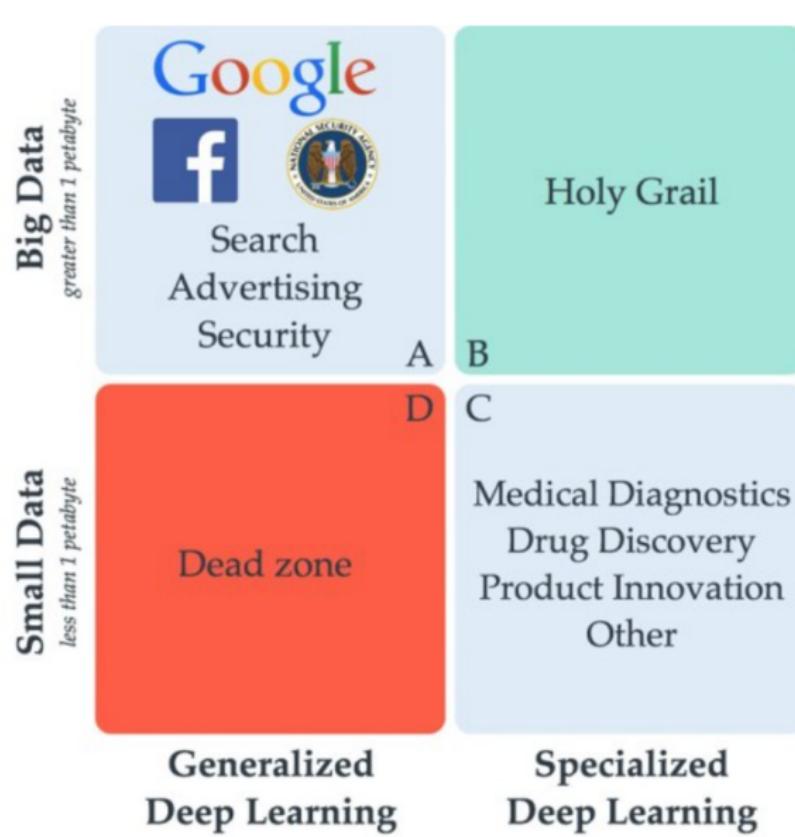
Perspectives

Perspectives: Data Integration

Integrating multi-omics, clinical and real world data



Perspectives: New Machine Learning Approaches



N. Sokolovska, H. T. Nguyen, K. Clément, J.-D. Zucker. Deep Self-Organising Maps for Efficient Heterogeneous Biomedical Signatures Extraction. International Joint Conference on Neural Networks (IJCNN), 2016.

Motivation

- ▶ Heterogeneous data integration for biomedical data
- ▶ Signature extraction (feature/biomarker selection)
- ▶ Convenient visualisation of results
- ▶ A hierarchical model (hopefully) can reveal dependencies between components to understand biological network structure

Self-Organising Maps for Feature Selection

Our framework relies on:

- ▶ Self-Organizing Maps
- ▶ Deep Learning
- ▶ Forward-Backward updates
- ▶ Heterogeneous Data Integration

Self-Organising Maps for Feature Selection

We propose a **hierarchical feature selection scheme with SOM**. The features in the backward step are drawn randomly.

```
// FORWARD
for each layer  $l \in L$  // bottom up do
    Run a SOM
    Select representatives from each cluster to propagate them to
    an upper level
end for
// BACKWARD
for each layer  $l \in L$  // top down do
    Estimate accuracy for level  $l$ 
    Greedily update selected features
end for
```

Signatures of Metabolic Health

- ▶ How to **stratify patients** in order to choose an efficient appropriate personalised medical treatment ?
- ▶ A quantitative metagenomic analysis stratified patients into two groups: **group with low gene gut flora count (LGC) and high gene gut flora count (HGC) group.**
- ▶ The LGC individuals have a higher insulin-resistance and low-grade inflammation, and therefore the gene richness is strongly associated with obesity-driven diseases.
- ▶ The individuals from a low gene count group seemed to have an increased risk to develop obesity-related cardiometabolic risk compared to the patients from the high gene count group.

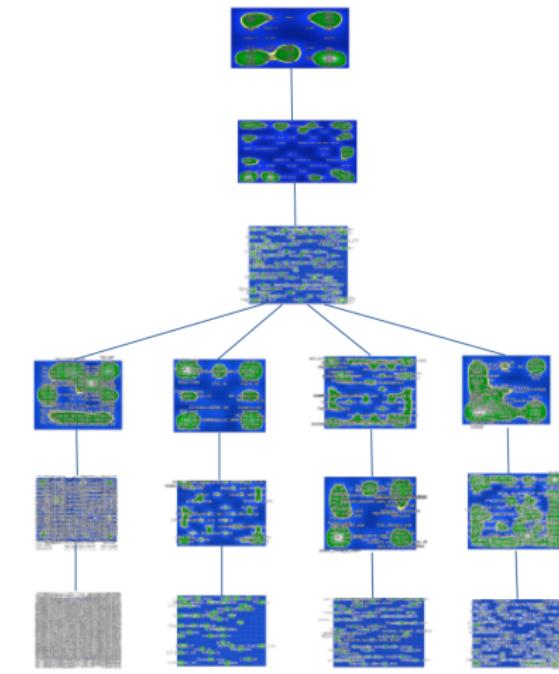
Experiments on MicroObese Cohort

- ▶ **Observations:** 49 patients have been hired and examined at the Pitié-Salpêtrière hospital, Paris, France ¹
- ▶ **Classes:** High Gene Count/Low Gene Count groups
- ▶ **Data integrated**
 - ▶ Meta-data (150)
 - ▶ Genes of adipose tissue (24,000)
 - ▶ Gut flora metagenomics (3 millions)



¹A. Cotillard and al. *Dietary intervention impact on gut microbial gene richness*. Nature, 500:585–588, 2013

Experiments on MicroObese Cohort



The hierarchy of SOM. For three lower levels, from left to right: MGS, environmental variables, host, and adipose tissue microarray data. Three upper layers perform data integration from four data sources.

Experiments on MicroObese Cohort



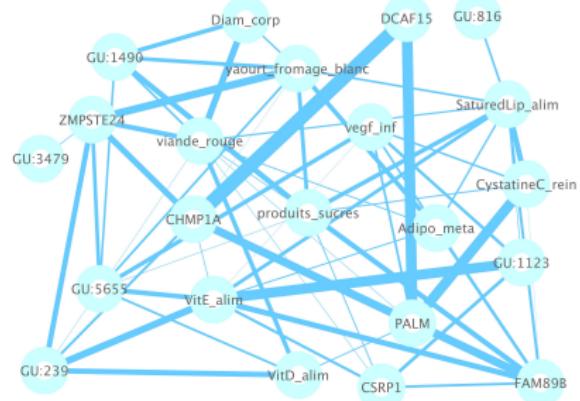
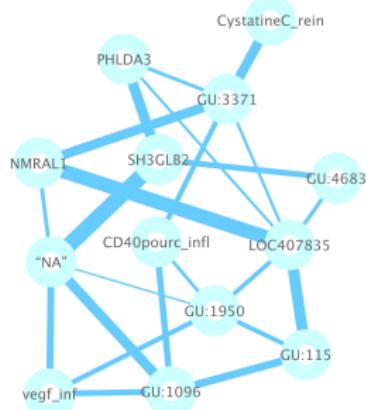
Signature of the high gene count group which is associated with a better health.

Experiments on MicroObese Cohort



Signature of the low gene count group which is associated with higher inflammation.

Experiments on MicroObese Cohort



On the left: Bayesian network of the selected features associated with the HGC, on the right: with LGC.

Experiments on MicroObese Cohort



A signature which discriminates high gene count and low gene count groups.

Perspectives and Conclusions

Conclusions

- ▶ The proposed deep method is efficient and competitive
- ▶ Data integration is easy
- ▶ Efficient dimensionality reduction

Perspectives

- ▶ Currently working on similar deep learning approaches
- ▶ Structure learning
- ▶ Theoretical foundations on deep learning

- ▶ N. Sokolovska, K. Clément, J.-D. Zucker. Deep Kernel Dimensionality Reduction for Scalable Data Integration. International Journal of Approximate Reasoning, 2016
- ▶ N. Sokolovska, S. Rizkalla, K. Clément, J.-D. Zucker. Continuous and Discrete Deep Classifiers for Data Integration. International Symposium on Intelligent Data Analysis (IDA), 2015

Motivation

- ▶ Heterogeneous bioclinical data
 - ▶ clinical and alimentary data
 - ▶ lipidomics (adipose tissue)
 - ▶ metagenomics (genes of gut flora)
- ▶ Human beings can be stratified into High Gene Count/Low Gene group (binary classification)
- ▶ Your (personalized) medical treatment depends on to what group you belong

Motivation

Data integration

- ▶ relates information of different nature
- ▶ hopefully increases performance of supervised learning
- ▶ increases dimensionality of a problem
- ▶ How to do learning with the integrated data efficiently?

Supervised Deep Kernel Dimensionality Reduction

KDR² method assumes

- ▶ there is a r -dimensional subspace ($r \ll d$) which is referred to as the effective subspace
- ▶ The dimensionality reduction can be viewed as a procedure testing conditional independence of variables such that

$$p(y|x) = \hat{p}(y|\theta^T x).$$

²K. Fukumizu, F. Bach, M. I. Jordan. *Kernel dimensionality reduction for supervised learning*. In NIPS, 2003.

Deep Dimensionality Reduction

We introduce a deep semiparametric model with D layers

$$p(y|x) = \hat{p}\left(y|\theta_D^T(\theta_{D-1}^T \dots (\underbrace{\theta_1^T(\theta_0^T x)}_{x'}))\right),$$
$$\underbrace{x''}_{\dots}$$

where x' , x'' , \dots , are new representations in the deep structure that are learned simultaneously in one optimization procedure.

Simultaneous Parameters Optimization

We clearly see that θ_{j+1} depends on θ_j for all $j \in \{1, \dots, D\}$, and optimization can not be done separately for each layer. By the **implicit function theorem**, applying the chain rule, for each θ_j , except for θ_0 , we have

$$\frac{\partial \ell(\theta)}{\partial \theta_j} = \frac{\partial \ell(\theta)}{\partial \theta_{j'}} \left(\frac{\partial \ell^2(\theta)}{\partial \theta_{j'}^2} \right)^{-1} \frac{\partial \ell^2(\theta)}{\partial \theta_j \partial \theta_{j'}},$$

where $\ell = \det \hat{\Sigma}_{YY|U}$.

KDR vs DKDR

Why is the Deep KDR efficient?

- ▶ real data are always noisy, and a “good” clustering or dimensionality reduction can significantly reduce the noise
- ▶ if features are tied into clusters of “high quality”, then it is easier to detect a signal from data, and therefore the generalizing classification accuracy is higher
- ▶ the hierarchical dimensionality reduction plays here a role of a filter, and a filter with multiple layers seems to perform better than a one-layer filter.

A Discrete Version of DKDR

A discrete classifier is easier to interpret than a continuous one

- ▶ Randomized rounding is a natural idea for rounding fractional solutions
- ▶ The algorithm starts from a real-valued solution, and for each parameter j of a model it draws a discrete solution from $\{-1, 0, 1\}$ according to $p(\theta_j)$
- ▶ where θ_j is a value of parameter j after some bound constrained optimization.

Experiments on Standard Data

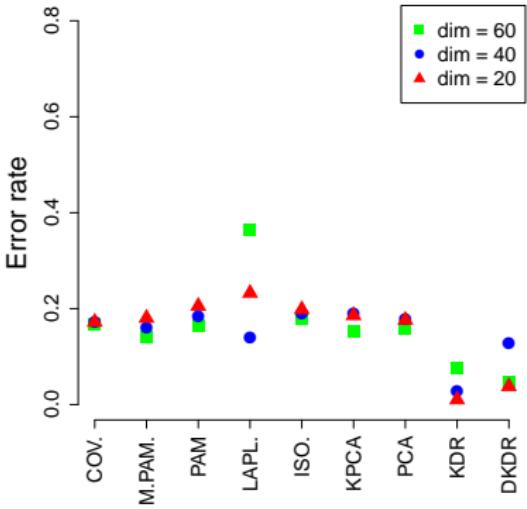
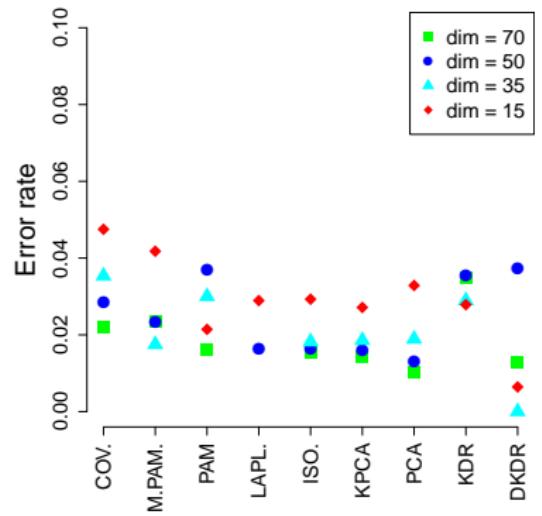
► Alon Data

- ▶ contains 62 patients and 2000 gene expressions (Affymetrix oligonucleotide array) of colon tissues. The patients are coming from two classes: 40 patients are diagnosed with a tumor, and 22 patients have normal colon tissues.

► Golub Data

- ▶ 72 patients and about 7000 gene expressions (Affymetrix probes). Among these patients, 47 subjects have acute lymphoblastic leukemia, and 25 are diagnosed with acute myeloid leukemia, therefore, we have a classification problem with 2 classes.

Experiments on Standard Data



Experiments on Golub Data Set (on the left) and Alon Data (on the right). Error rate as a function of dimensionality reduction method and dimension in reduced models.

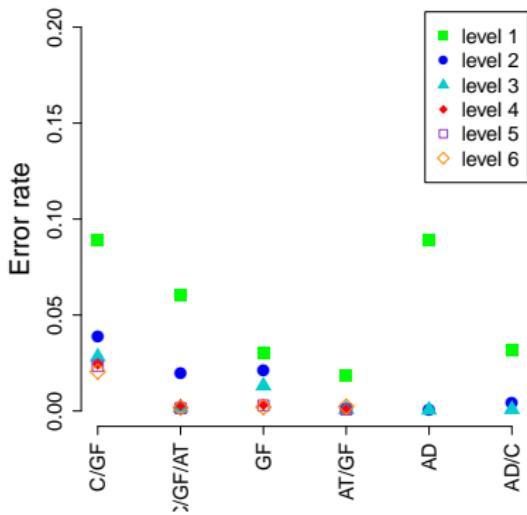
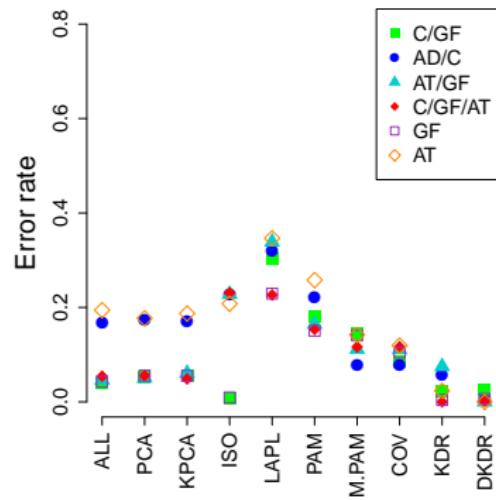
Experiments on MicroObese Cohort

- ▶ Observations: 49 patients have been hired and examined at the Pitié-Salpêtrière hospital, Paris, France ³
- ▶ Classes: High Gene Count/Low Gene Count groups
- ▶ Data integrated
 - ▶ Meta-data (150)
 - ▶ Genes of adipose tissue (24,000)
 - ▶ Gut flora metagenomics (3 millions)



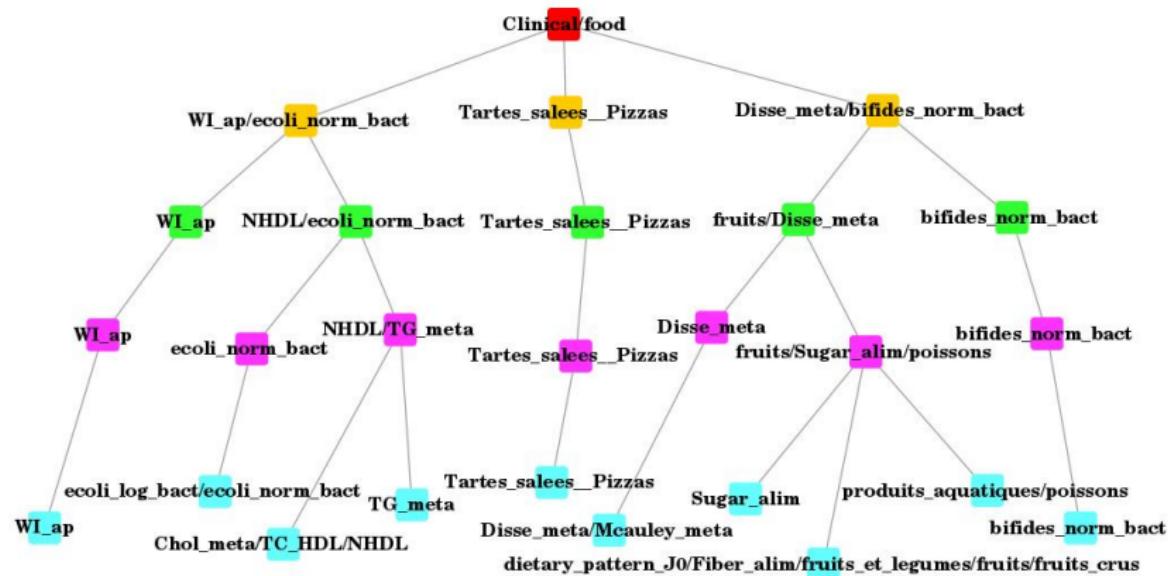
³A. Cotillard and al. *Dietary intervention impact on gut microbial gene richness*. Nature, 500:585?588, 2013

Experiments on MicroObese Cohort

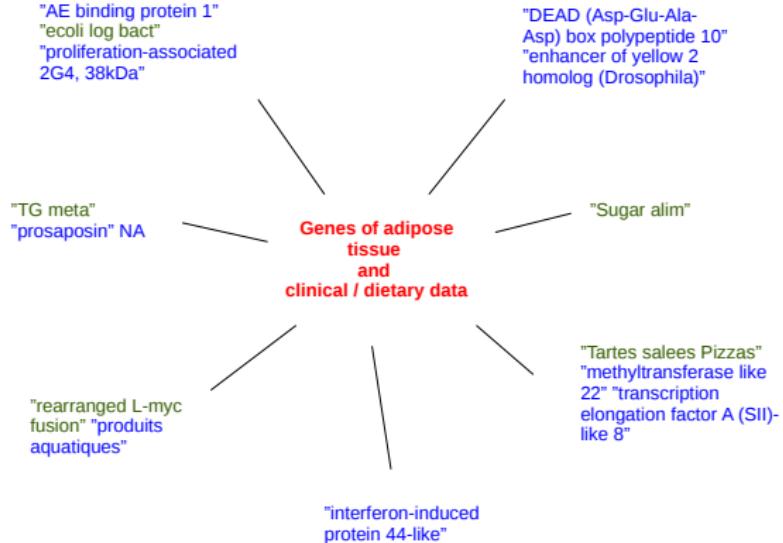


MicroObese Cohort. On the left: error rate as a function of dimensionality reduction method and data integrated into the model. On the right: error rate as a function of data integrated and level in the hierarchy.

A hierarchy of clinical parameters of MicrObese data



A signature of HGC/LGC groups based on clinical parameters and adipose tissue genes



Perspectives and Conclusions

Conclusions

- ▶ The proposed deep method is efficient and competitive
- ▶ Data integration is easy
- ▶ Efficient dimensionality reduction

Perspectives

- ▶ Currently working on similar deep learning approaches
- ▶ Structure learning
- ▶ Theoretical foundations on deep learning