SCIENTIFIC REPORTS
nature research

DeePathology: Deep Multi-Task Learning for Inferring Molecular Pathology from Cancer Transcriptome

Behrooz Azarkhalili[1,4], Ali Saberi[2], Hamidreza Chitsaz[3] & Ali Sharifi-Zarchi[2*]

BMED 8813 BHI Presenter: **G-6**

Seonggeon Cho, Rohan Bhukar, Bryce Butler, Zhonghao Dai




Spring, 2021

Current Problems

- Improving the accuracy of *cancer diagnostics* is important in diagnostic results and its subsequent treatment strategies. Despite improvements over the years significant *diagnostic errors* still occur.

- Current *multiclass algorithms* geared towards diagnosis of the whole-transcriptomes of tissue biopsies are still not fully developed

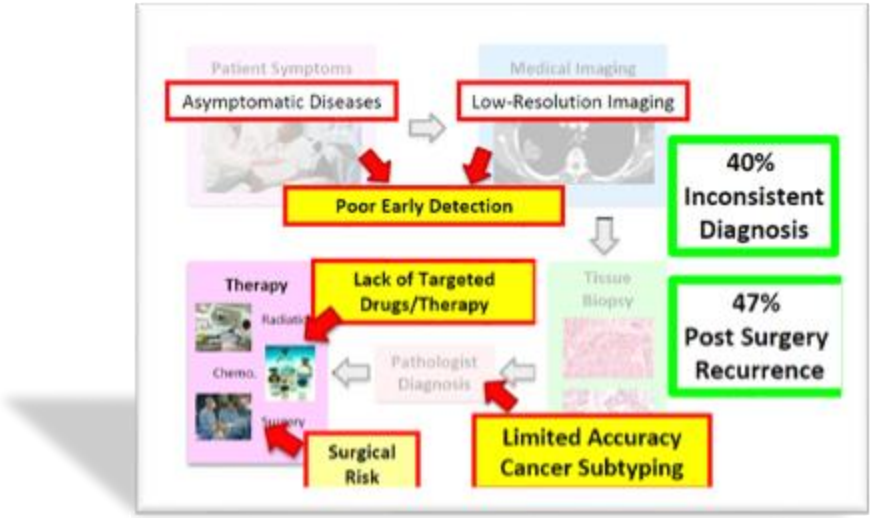


Figure 1

Source : Dr. Wang, 1st week Lecture PPT - BMED 8813

Limitations of current cancer diagnostics

- Current Molecular pathology methods can only detect a limited number of genes or proteins
- The approach to using the whole transcriptome of tissue biopsies have proven to be computationally challenging. Multiclass algorithms, like Optimal Feature Weighting, employed so far have not been successful.
- The *inaccuracy* of cancer classification

Goal

Develop accurate diagnostic method using Deep Neural Networks in a *multi-task learning approach* to infer different biological and clinical information from *whole-transcriptome of tissue biopsies*

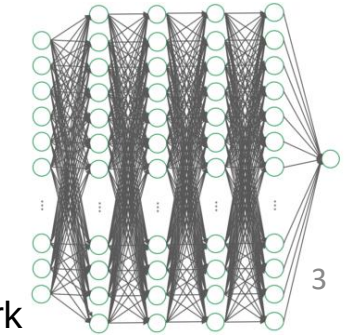


Figure 1 : A dense neural network

Workflow towards solution

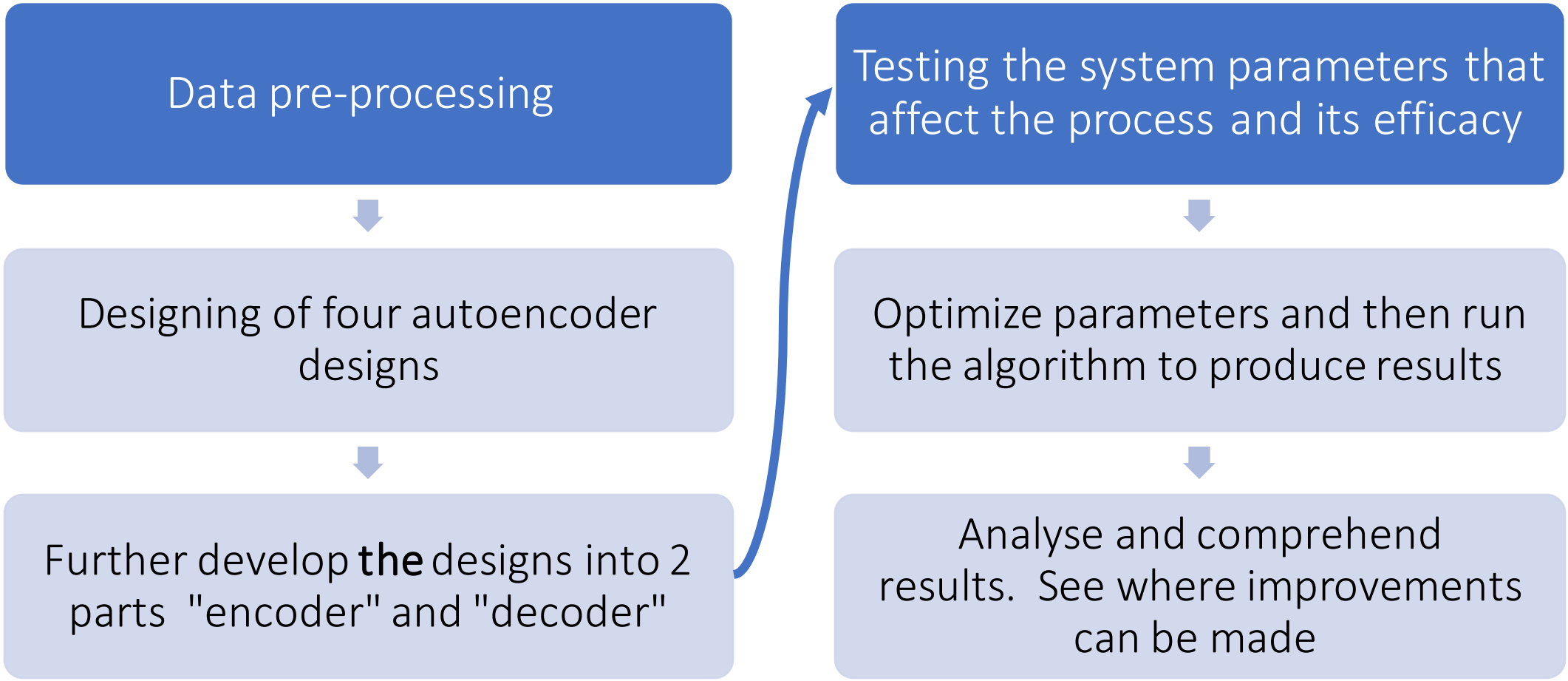


Chart 1. system workflow diagram

Graphical illustration of one DNN architecture of the DeePathology

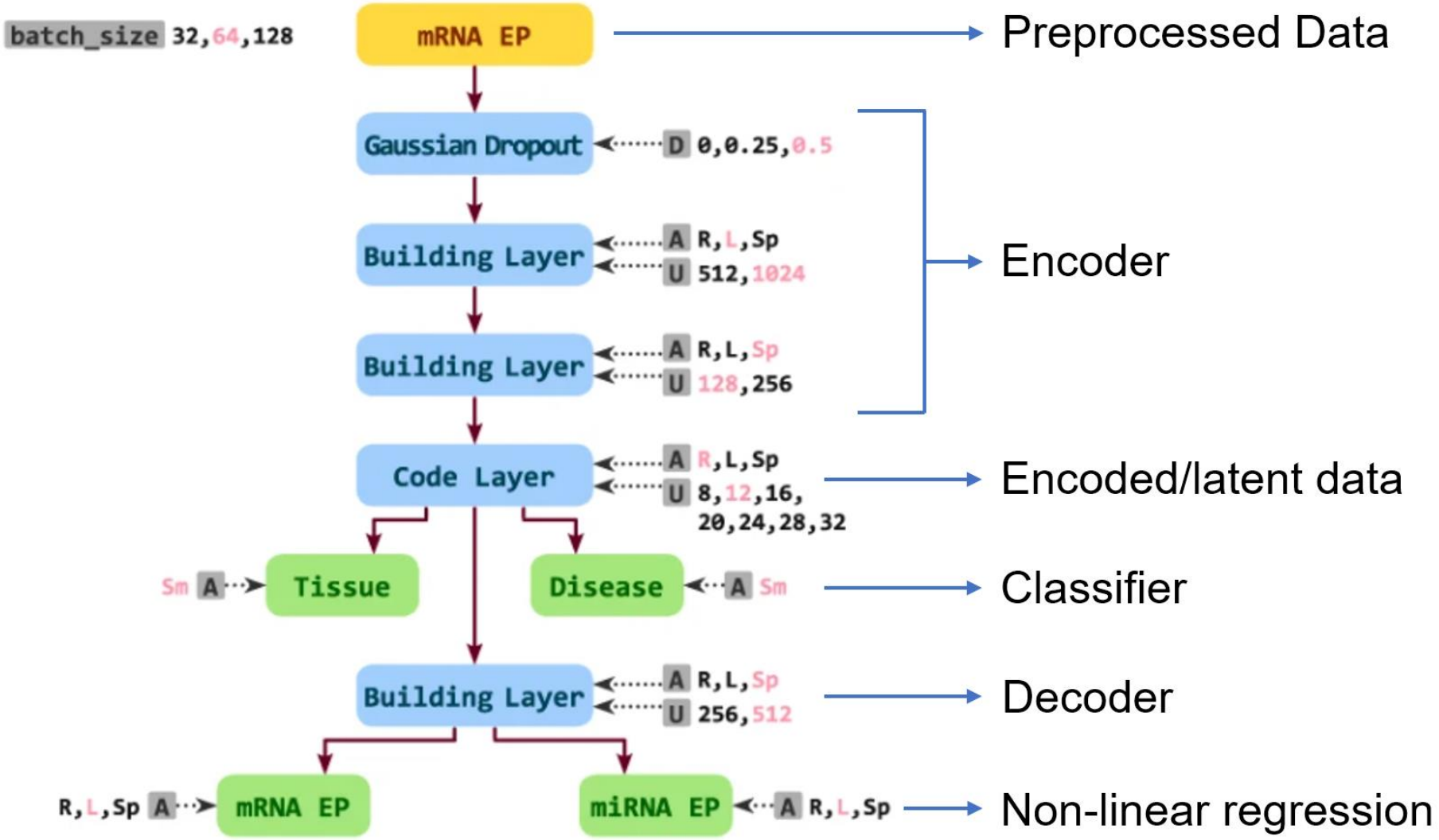
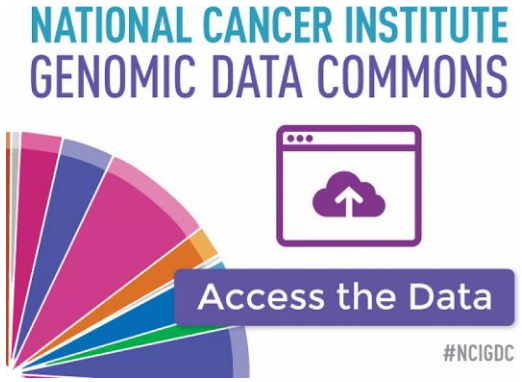


Figure 1: DNN architecture

Input data

- Data source: Genomic Data Commons
  - Cancer Genome Atlas (**TCGA**)
  - Therapeutically Applicable Research to Generate Effective Treatments (**TARGET**)
- Samples with both mRNA and miRNA expression profiles available were kept. **10,150 tumor** and **637 normal samples** across various tissue types.
- **90%** of data used as ***training dataset***, **10%** used as ***testing dataset***.



	GDC Data		
Tissue	**Tumor**	**Normal**	**Total**
Adrenal Gland	262	3	265
Bile Duct	36	9	45
Bladder	411	19	430
Blood	126	0	126
Bone Marrow	83	0	83
Brain	525	0	525
Breast	1088	104	1192
Cervix	306	3	309
Colorectal	611	12	623
Esophagus	162	11	173
Eye	80	0	80
Head and Neck	493	44	537
Kidney	964	141	1105
Liver	370	50	420
Lung	987	59	1046
Lymph Nodes	47	0	47
Ovary	376	0	376
Pancreas	178	4	182
Pleura	86	0	86
Prostate	495	52	547
Skin	449	1	450
Soft Tissue	261	0	261
Stomach	372	32	404
Testis	156	0	156
Thymus	119	2	121
Thyroid	509	58	567
Uterus	598	33	631
All	**10150**	**637**	**10787**

Table 1: The number of samples used for each tissue type

Information extraction from mRNA data

mRNA expression profile size: **19,671** → latent vector of size: **8**

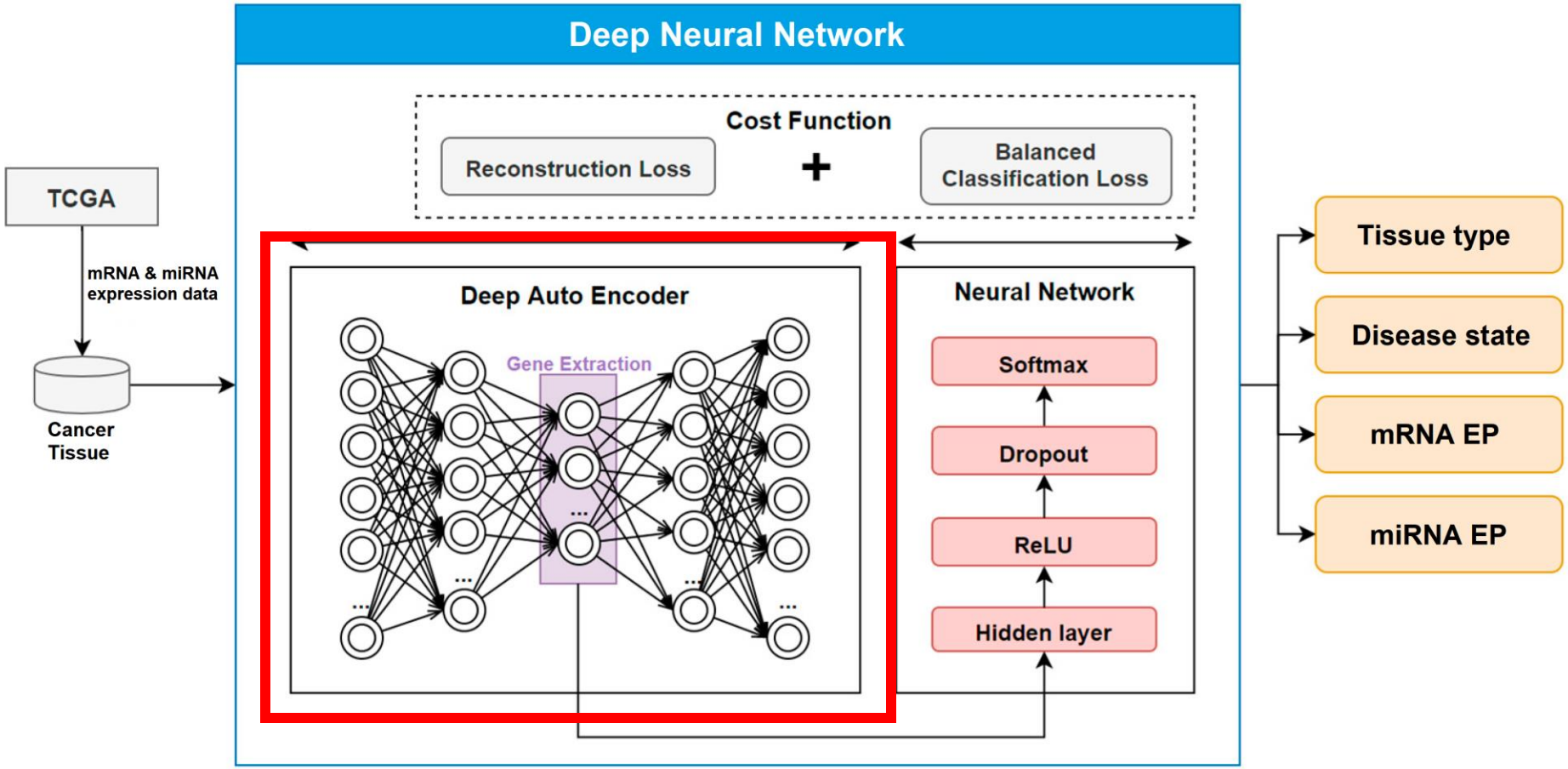


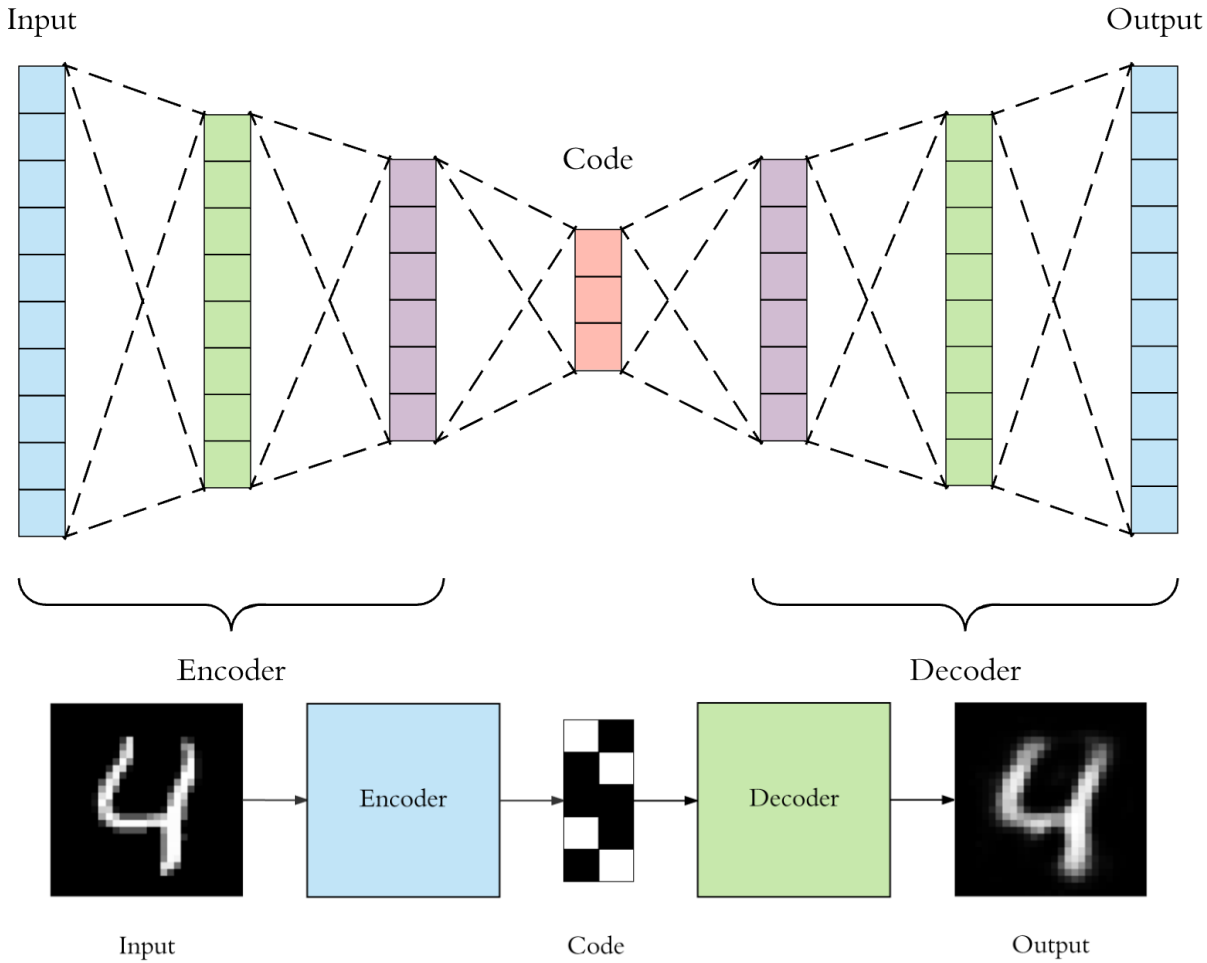
Figure 1: Overview of the Deep Neural Network

What is deep auto encoder?

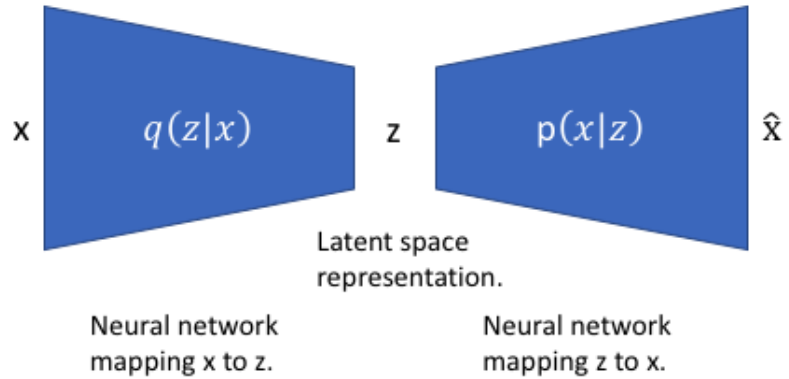
Compresses input into a latent space representation, and then reconstructs the input back from this representation

Application

- Data compression
- Dimension reduction
- Manifold learning
- Feature learning



Encoder



Contractive autoencoder:

$$\mathcal{J} = \sum_{x \in D} L(x, \tilde{x}(\phi, \psi, x)) + \lambda \left|\left| J_{\psi}(x) \right|\right|_F^2$$

- Eq 1

Loss term

Regularization term

Variation autoencoder:

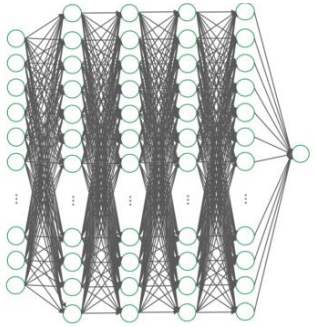
$$\mathcal{J}(\phi, \theta, x) = D_{KL}\left(q_{\phi}(z|x) \middle|\middle| p_{\theta}(z)\right) - \mathbb{E}_{q_{\phi}((z|x))}(\log(p_{\theta}(x|z)))$$

- Eq 2

Stay close to true prior distribution

Reconstruction loss

Understanding Dropout and Gaussian dropouts in Neural Networks



A dense neural network.
Figure 1.

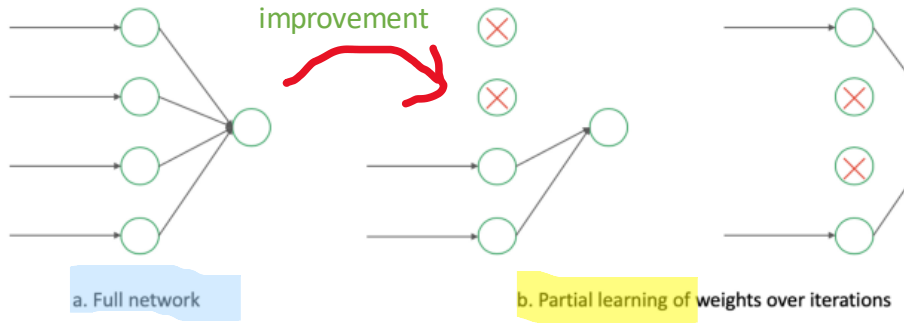
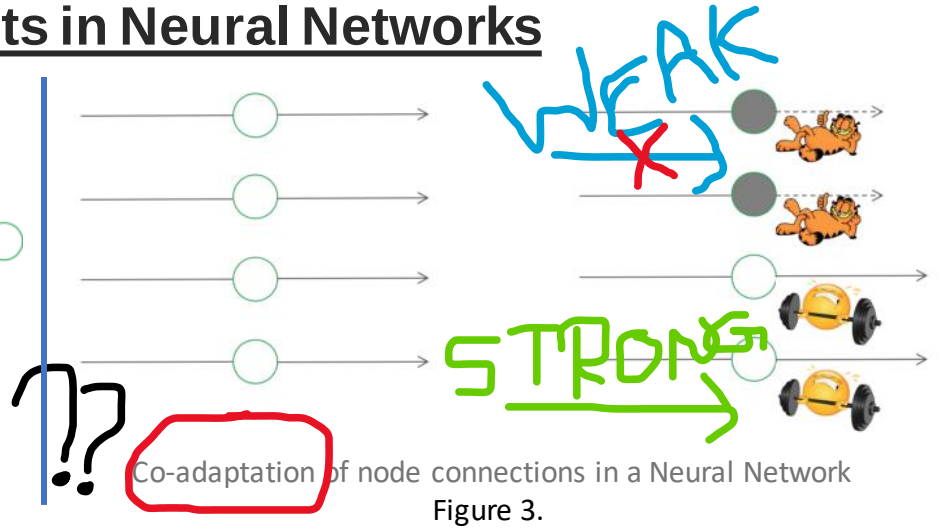


Illustration of learning a part of the network in each iteration.
Figure 2.



Co-adaptation of node connections in a Neural Network
Figure 3.

Srivastava et al. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. JMLR 06, 2014

Dropout equivalent to regularized Network

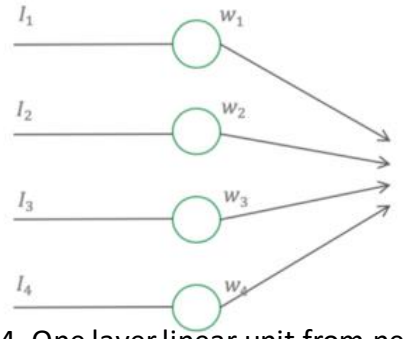


Figure 4. One layer linear unit from network

OLS loss function

loss for a regular network

$$E_N = \frac{1}{2}\left(t - \sum_{i=1}^{n} w_i' I_i\right)^2 \quad \text{- Eq 1}$$

$$E_D = \frac{1}{2}\left(t - \sum_{i=1}^{n} \delta_i w_i I_i\right)^2 \quad \text{- Eq 2}$$

loss for a dropout network

upon derivation

$$E_R = \frac{1}{2}\left(t - \sum_{i=1}^{n} p_i w_i I_i\right)^2 + \sum_{i=1}^{n} p_i(1-p_i) w_i^2 I_i^2 \quad \text{- Eq 3}$$

dropout loss minimization is equivalent to a regularized network minimization

What is Gaussian - Dropout?

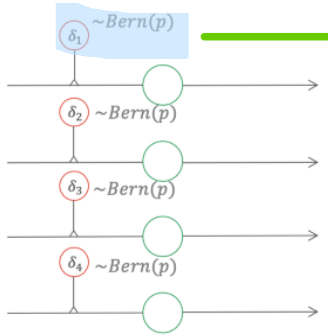


Figure 5. Dropout as Bernoulli gate on connections

Bernoulli gate replaced by Gaussian gate

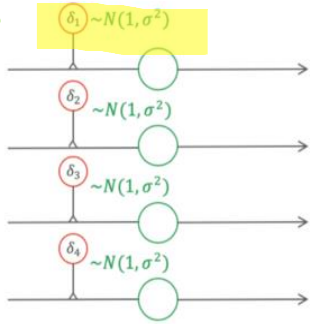


Figure 6. Dropout generalized to Gaussian gate

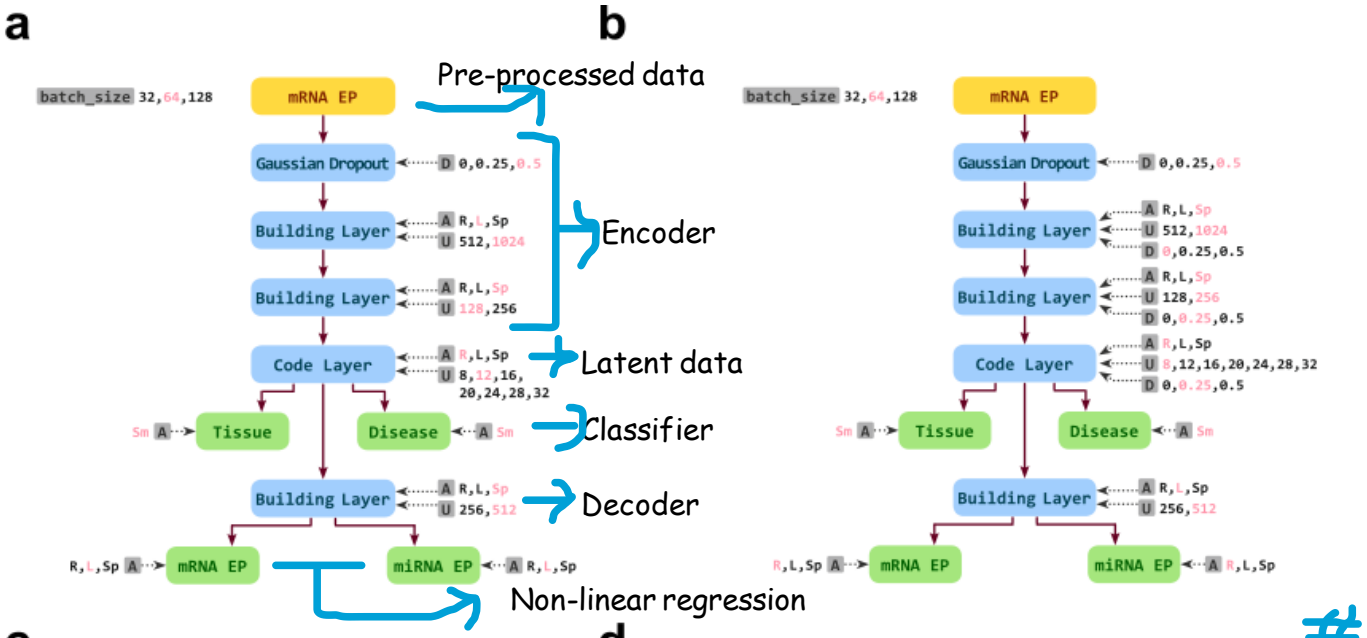
expected value of the activation remains same

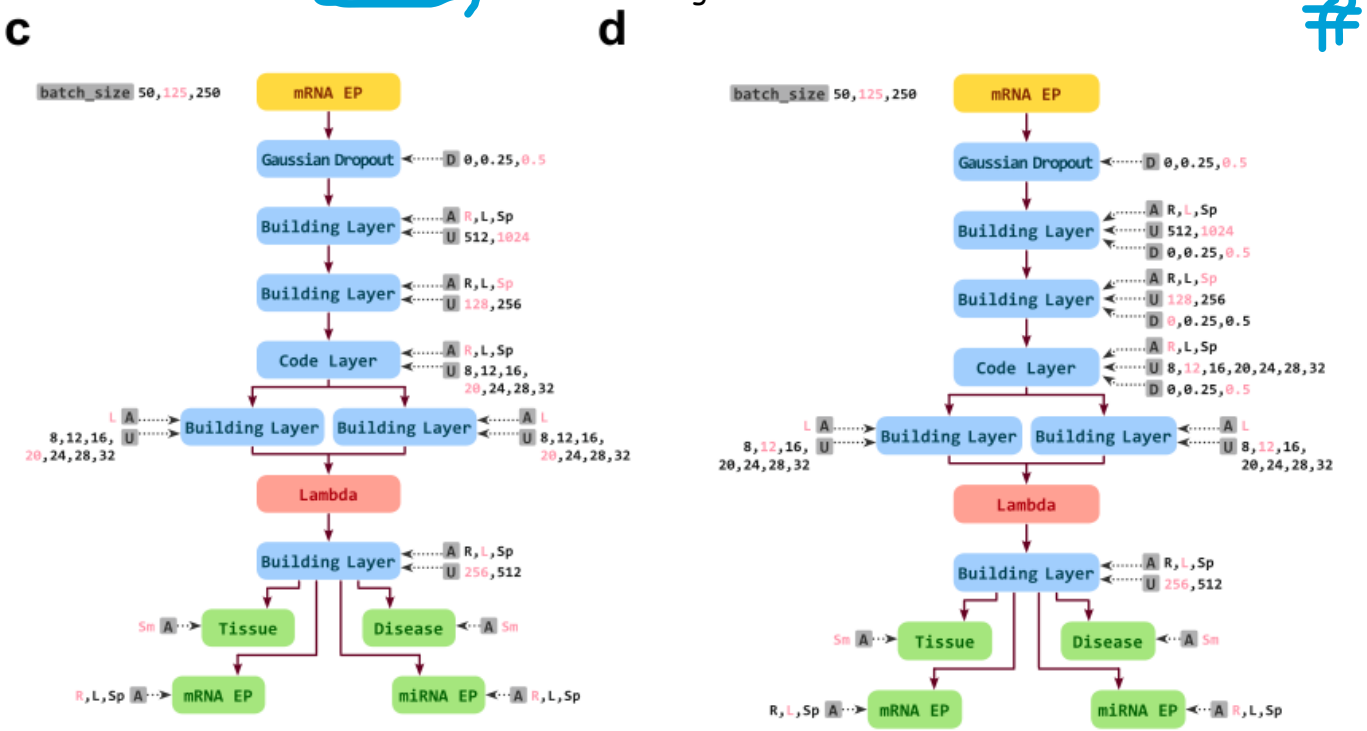
$$E\left[\sum \delta_i w_i I_i\right] = \sum w_i I_i, \quad \text{if } \delta_i \sim N(1, \sigma_i)$$

- Eq 4

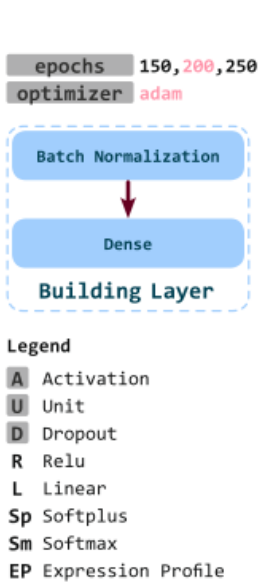
resulting in computational performance advantage

Different architectures of DNN used for this study.





#



Their problem has :-

- 2 *regression* tasks
  a) reproduce mRNA EP from latent variables
  b) reproduce miRNA profiles.
- 2 *classification* tasks
  a) find tissue state of each sample
  b) find disease state.

expression profiles normalization (by max-norm : Scikit)

constructed multi-input and multi-output models (Keras)

With 4 different Neural Network architectures :

a) **Variational autoencoder** (VAE)
b) **Dropout-VAE**, an extension of VAE with Bernoulli & Gaussian dropout layers for denoising
c) **Contractive autoencoder** (CAE)
d) **Dropout-CAE**, the extended CAE with Bernoulli & Gaussian dropout layers

NOTE :

- in each network, layers = boxes & connections = arrows;
- evaluated hyperparameters of each layer are next to each layer;
- hyperparameter values in red show optimal parameters (by hyperparameter optimization)

Balanced Accuracy

- for classification tasks, *Cosine similarity* was used as the loss function (due to ease of use in hyperparameter optimization)
- for regression tasks, *mean squared error* (MSE) was used as loss function
- **TOTAL LOSS** = weighted sum of individual losses (w/ different wts. for different tasks)

- since for imbalanced datasets (which the authors had), conventional accuracy is misleading => used **BALANCED ACCURACY**

Adjusted sample weights → $\hat{w}_i = \frac{w_i}{\sum_j 1(y_i = y_j) w_j}$ ← weight of sample i

$$\text{Balanced accuracy}\,(y, \hat{y}, w) = \frac{1}{\sum \hat{w}_i} \sum_i 1(\hat{y}_i = y_i)\hat{w}_i$$

true label

predicted label

where $1(x)$ is the indicator function.

Batch Normalization

- training DNNs have a major issue : *altered distribution* of every layer inputs (weights & parameters of back layers change)
- Above is called *internal covariate shift*, which causes lowering of learning rate
- **BATCH NORMALIZATION** (BN) is the solution to address this.
  - standardizes inputs of each layer => resulting in mean and standard deviation of 0 and 1.
  - scaling these values by a linear function
  - extending BN to hidden layers => improves training speed

- Each building layer of their Network, consists of BN layer followed by a dense layer.

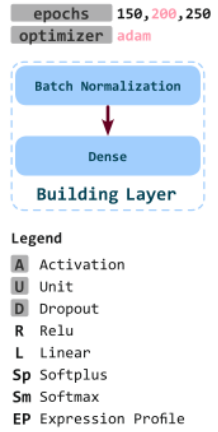


Figure 1. Building layer from the DNN architectures

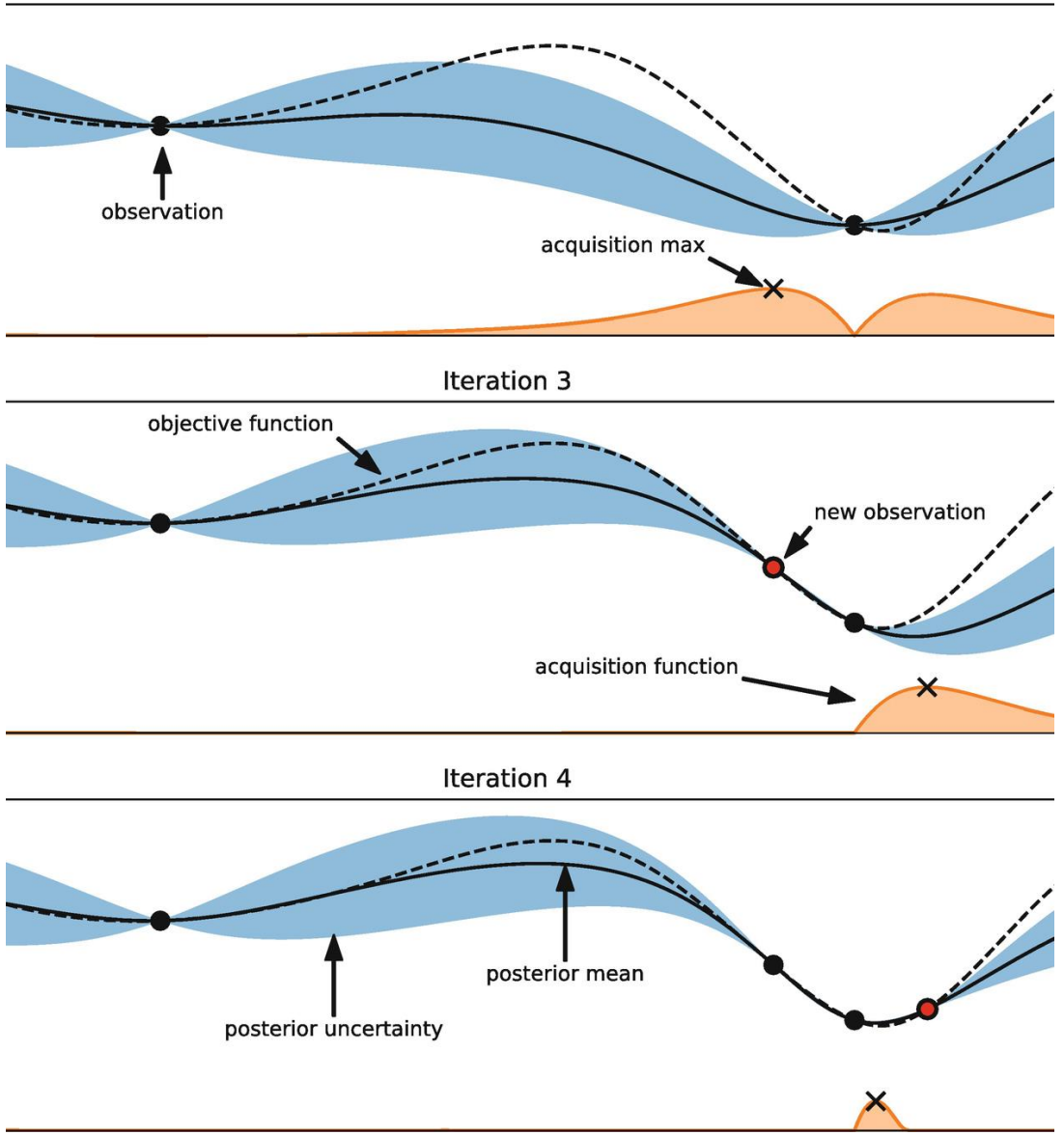
Decoder

- Non-linear regression
  - Reproducing mRNA and miRNA EP as one of the outputs that is as close to the original input as possible.

- Classification
  - Predicting the sample tissue of origin, among 27 different tissues
  - Predicting the sample disease state (normal or one of 33 different cancer types)

Hyperparameter Optimization/Tunning

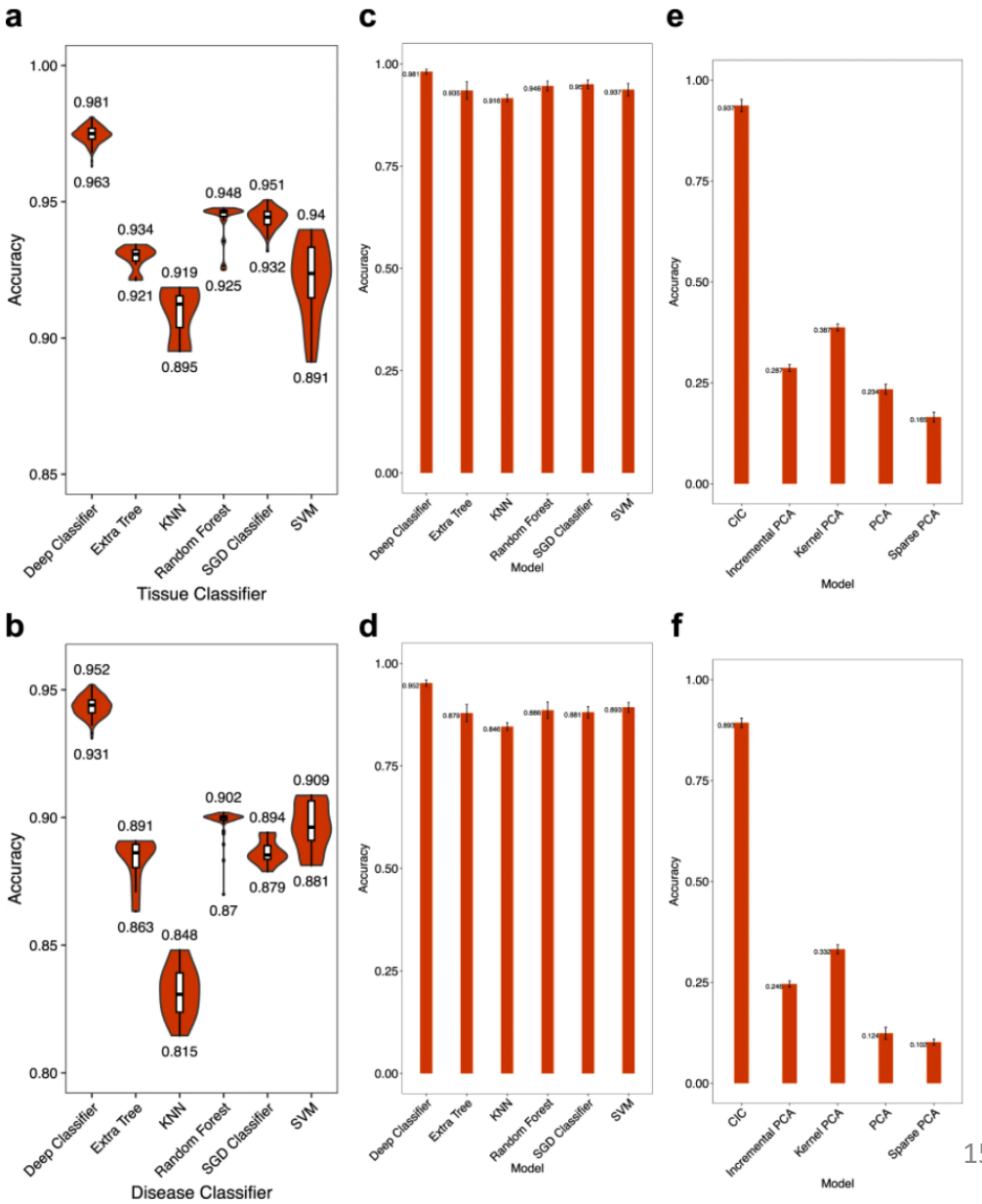
- Hyperparameters are different parameter values that control the learning process and affects the model's performance
- To optimize these parameters, strategies like grid search and random search are employed. This model uses Hyperopt.
- This is a form of Bayesian optimization for parameter tuning that allows you to get the best parameters for a given model.
- Surrogate model, P(s | h),
  where s and h are the objective function score and hyperparameters.



Feurer M., Hutter F. (2019)
Hyperparameter Optimization

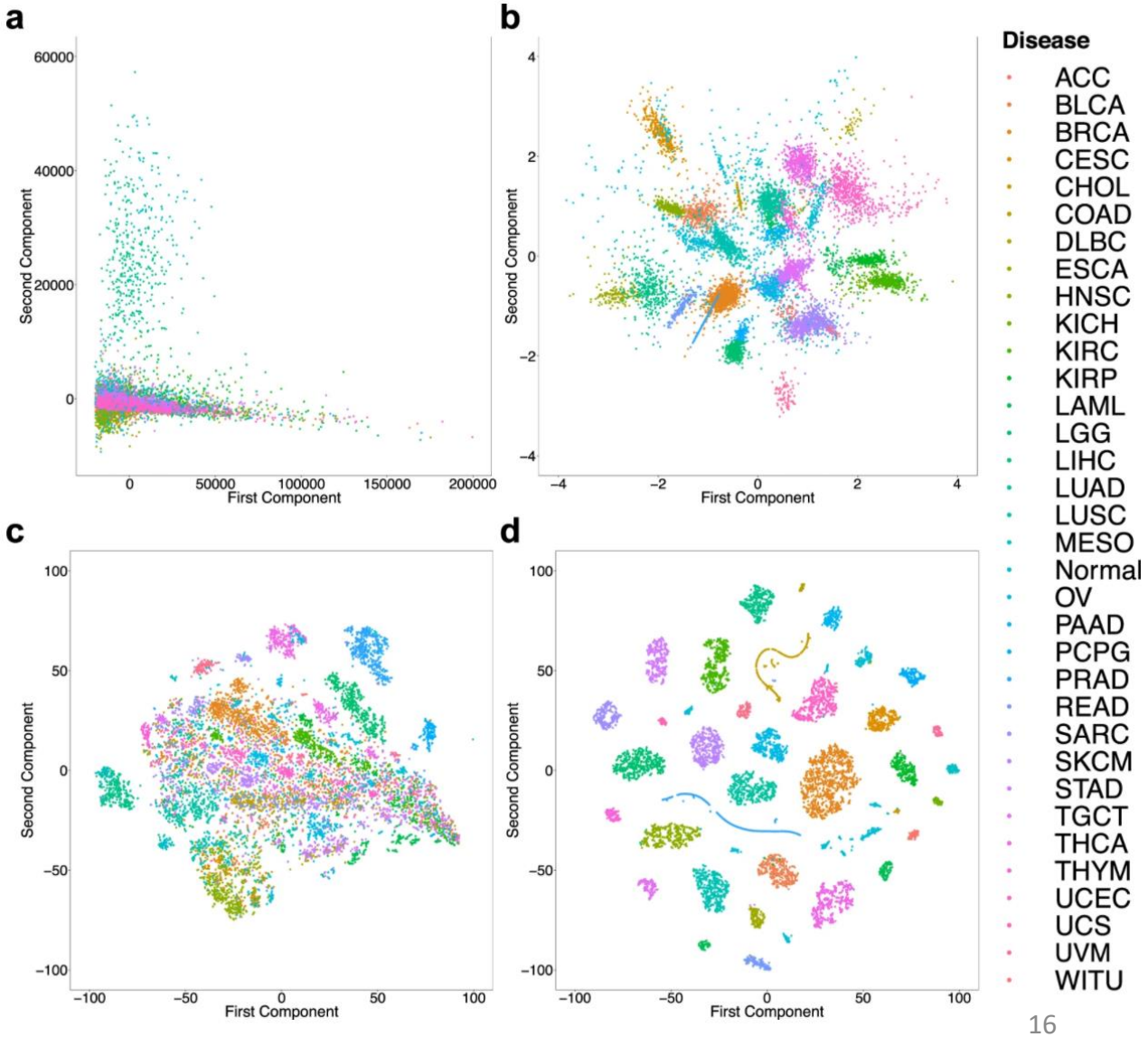
Performance evaluation of DNN

- DNN reaches balanced accuracy of 98.1% for tissue and 95.2% for disease classification, while that of other methods is at most 95.1% and 90.9% for tissue and disease classification
- The accuracy of the Deep classifier was better than other algorithms when provided with the same 8-dimensional Cell Identity Codes.
- CIC is a better dimensionality reduction than those obtained by classical algorithms



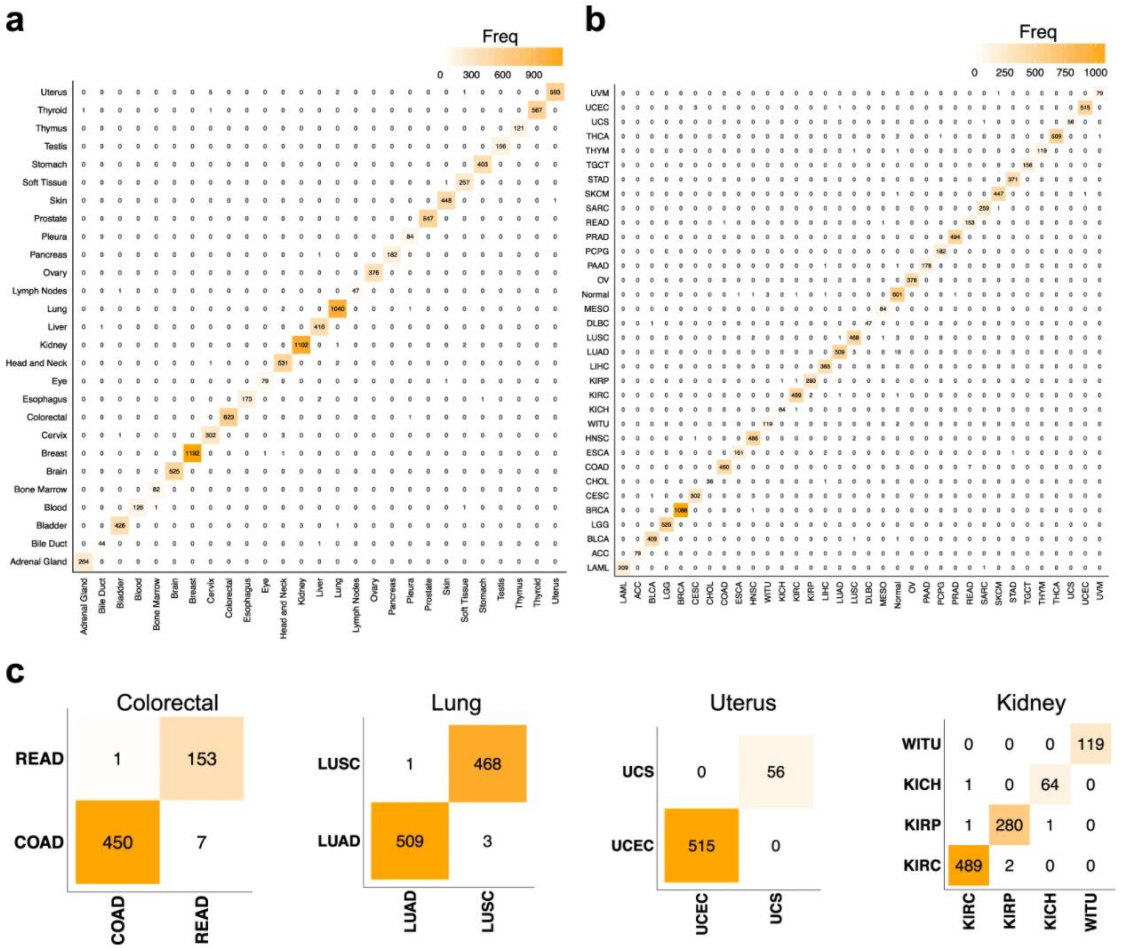
CIC versus mRNA profiles

- 19,671-dimensional mRNA expression profiles and the 8-dimensional CIC profiles are visualized using PCA and t-SNE.
- Original mRNA expression profiles from most of the studies overlap and impossible to discriminate disease states of the samples, while 8-dimensional CIC space is significantly better discriminative of the disease state.



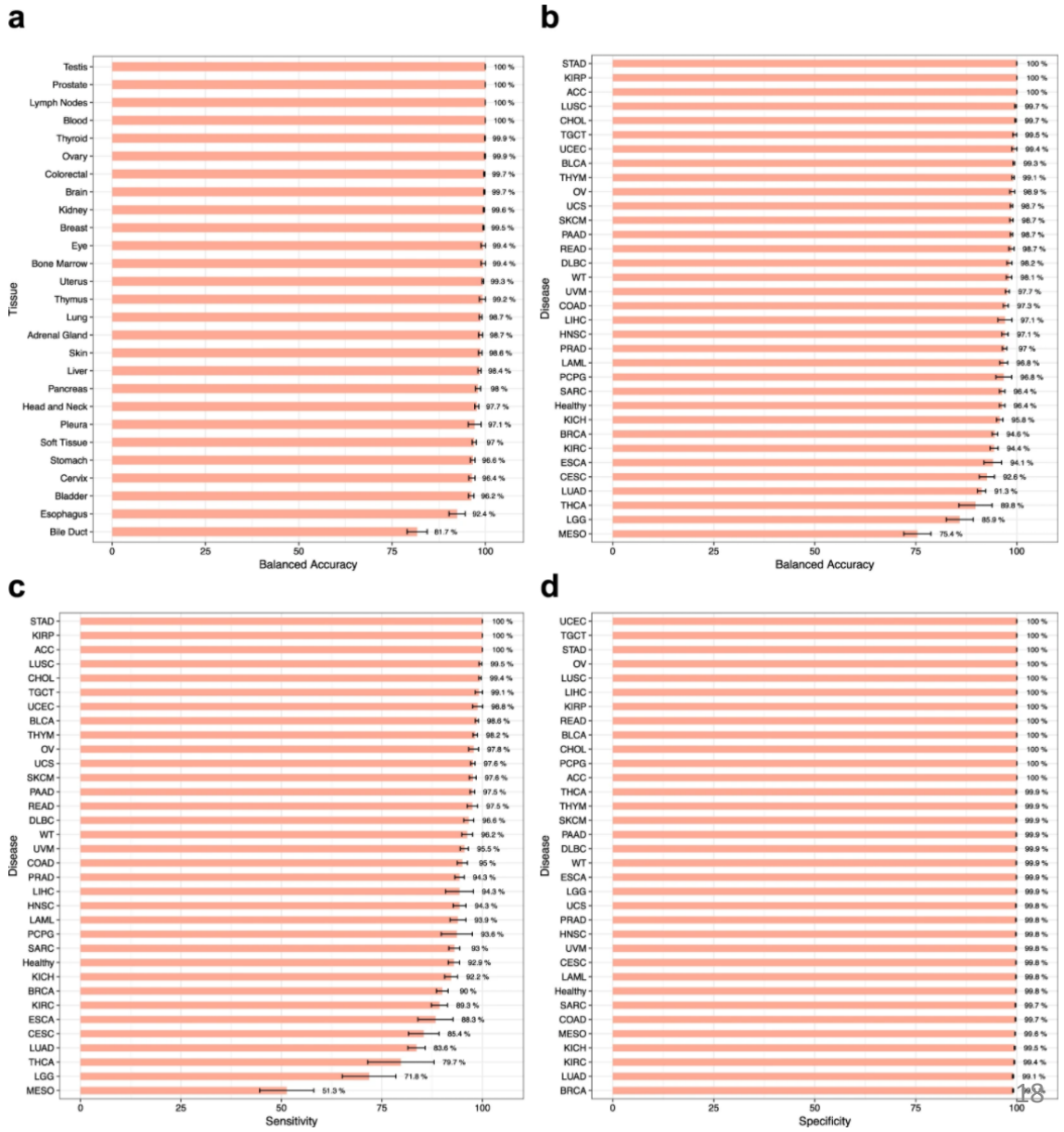
DNN accurately discriminates different cancer subtypes

- Confusion matrices for the tissue types and disease states shows over 99.5% *accuracy*.
- Confusion matrices for tissue types with more than one cancer subtypes retains over 99% *classification accuracy*

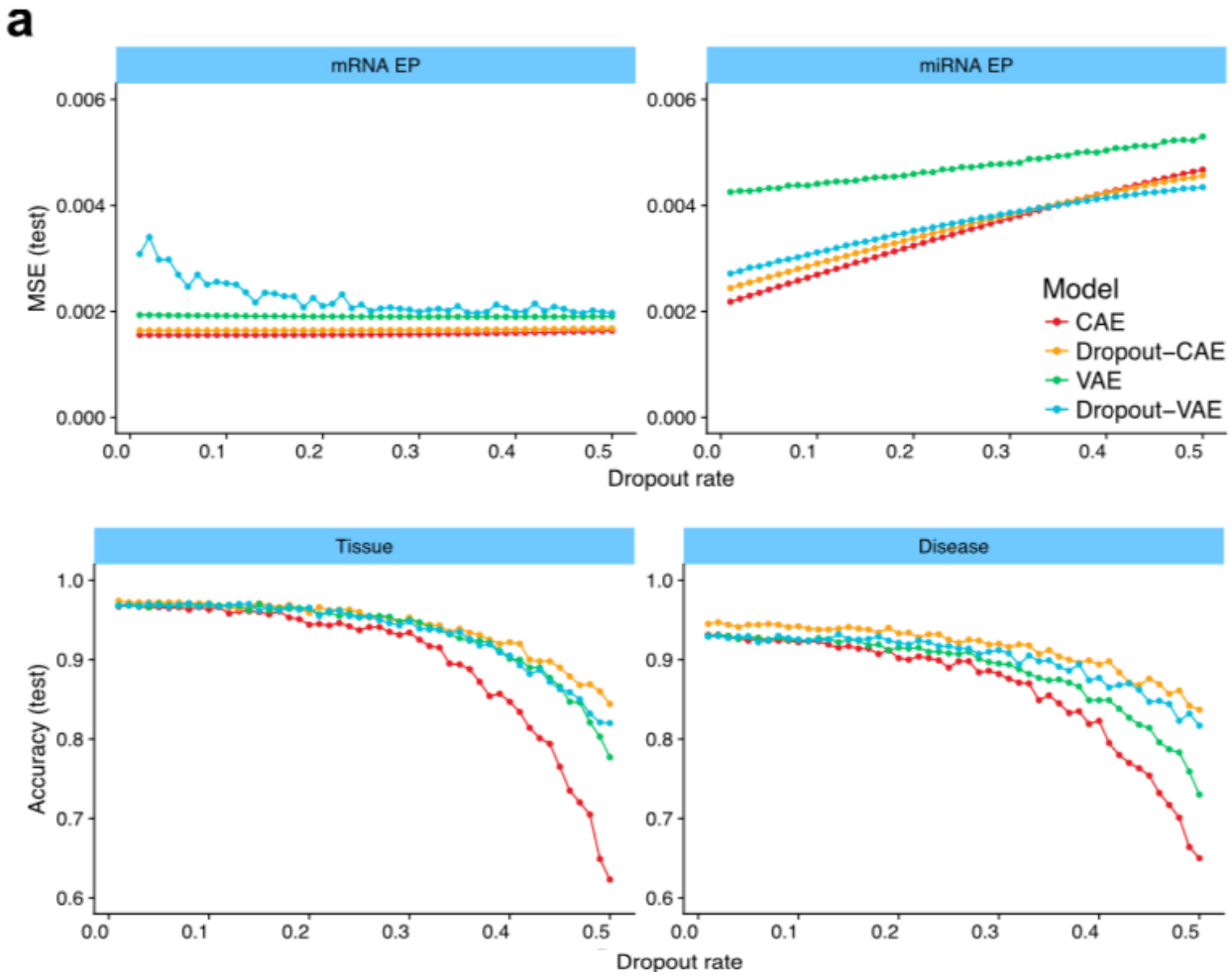


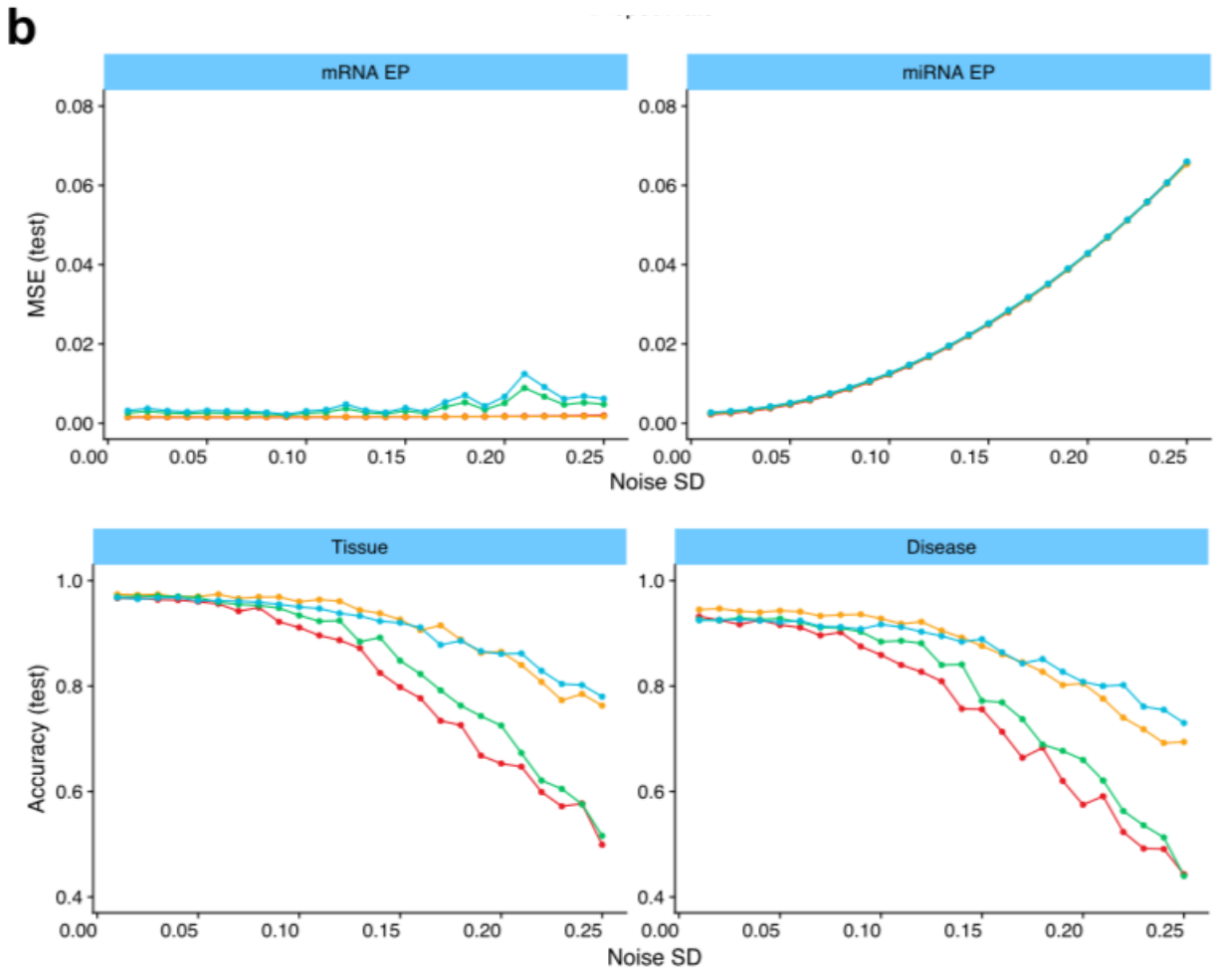
Stable classification of different tissues and cancer types by DNN

- *Balanced accuracy* of tissue classification: ≥99% for 14 tissues, and ≥95% for 25 out of 27 tissues. Disease classification: ≥99% accurate for 9 cancer types, and ≥95% for 25 cancer types and normal tissues
- *Low sensitivity* for certain disease classification maybe due to quality of "Normal" samples, since most of them are obtained from the tissues adjacent to cancer tumors



Resistance of DNN designs against missing values and noises





Summary and Conclusion

- They built deep neural network that can classify tissue type and cancer type from whole transcriptome of tissue biopsies
- They implemented contractive and variation autoencoder
- Their dropout and hyperparameter make the model robust against missing values and noisy input
- Accuracy of their classification method is higher than other classification method
- CIC, the latent vector created, is very light, accurate, and precise in classification and clustering.

Future Works

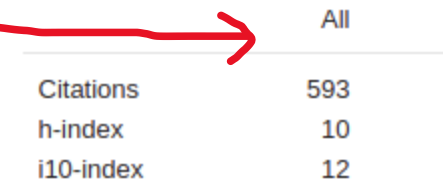
Their method can be used for finding novel biomarkers and regulatory interactions of different tissues and disease stats

***Strength* and *Weakness* of Paper**

Pros	Cons
Their algorithm has better accuracy in classifying tissue and cancer type compared to other classification tools.	Sampling is from tissue biopsy that limits application on early diagnosis
CIC provides precise and better quality of visualization.	For DNN to properly operate, large amount of data is required for training.
Their DNN provides accurate and sensitive method of classifying tissue type and cancer type	True normal sample is required for certain cancer type to increase accuracy in classification
They made robust DNN model that is resistant against noise and missing values	

Table of Summary

Who	• Behrooz Azarkhalili, Ali Saberi, Hamidreza Chitsaz, **Ali Sharifi-Zarchi*** • Sharif-Royan Bioinformatics Lab at the Department of Computer Engineering, SUT, Tehran; with interests and expertise in algorithm development & Deep Learning in Bioinformatics.
Where	• Sharif University of Technology: Tehran*, Tehran, IRAN • Royan Institute: Tehran, Tehran, IRAN
When	Study was published in 2019 in the Scientific Reports Journal in November edition. (Article number: 16526). Total citations so far = 4
What	• They built deep neural network that can classify tissue type and cancer type from whole transcriptome of tissue biopsies • A DNN (through multi-task & transfer learning) capable of inferencing many properties of biological-samples SIMULTANEOUSLY.
Why	• To Improve the accuracy of *cancer diagnostics* • Highlight applications of A.I. in molecular cancer pathology
How	• They built deep neural networks with contractive and variation autoencoders. • Hyperparameter was implemented to optimize models, and dropout method was used to create robust models



Feasibility and real world applications

- It is commonly known to ML experts that due to the black-box nature of deep neural networks, it is ***difficult to extract easy to under-stand patterns*** from the model and discover causal relationships;
- Many scientists have been making **efforts to utilize statistical methods to extract such hidden information** from the trained models;
- While **there exists methods to extract data** based on gradient variation that are categorized as sample-based and model-based approaches (Shap 48,49 , DeepLift 50 , LIME 51 , and Interpret 52), **such methods are still in development**;
- It might work **better in detecting patient sample tissues with > 1 type of cancer**, due to very high accuracies in detecting cancer-subtypes, and potentially pave the path for clinical utility in cancer-pathology;
- Addition of dropout layers, makes their system very **robust against missing data and noise**, which is an added improvement over previously existing solutions.

Learnings / takings from this article

- **Molecular cancer pathology is still a challenging area** in biomedical field where current solutions still have inaccuracies in sample detection apart from being highly expensive.
- Developing **low-cost solutions that leverage computationally efficient designs** (such as this proposed solution) of cancer sample detection is a need of the hour.
- **Deep NNs architectures** suggested with the article, provide significant results that **outperform classical ML** approaches in cancer type detection.
- Adding **batch normalization layers** to DNNs improves training speed of networks.
- A more advanced hyperparameter tuning Bayesian algorithm, utilizes **surrogate probability model**.
- Adding **dropout-layers to NN architectures makes them highly robust** against noise and missing values.

Motivations

- **With the data scarcity trend rapidly changing,** more and more data from biological samples is increasing everyday. Working towards using this proposed solutions in other cancer subtypes detection is one of the directions to explore.
- Learning a new type of DNN architecture with applications to biomedical sciences, sparks an idea to experiment this with Covid-sub phenotyping project.
- With applied A.I. methods having better results than current approaches such as Cox-PH, its exciting times to learn new things in a field with intersection on A.I. applications and Biomedical Science.

Additional questions?