Deep learning: a practical solution for real-world cases in medicine?

Ruxandra Stoean

University of Craiova

Group for Integrated Systems Engineering, University of Malaga

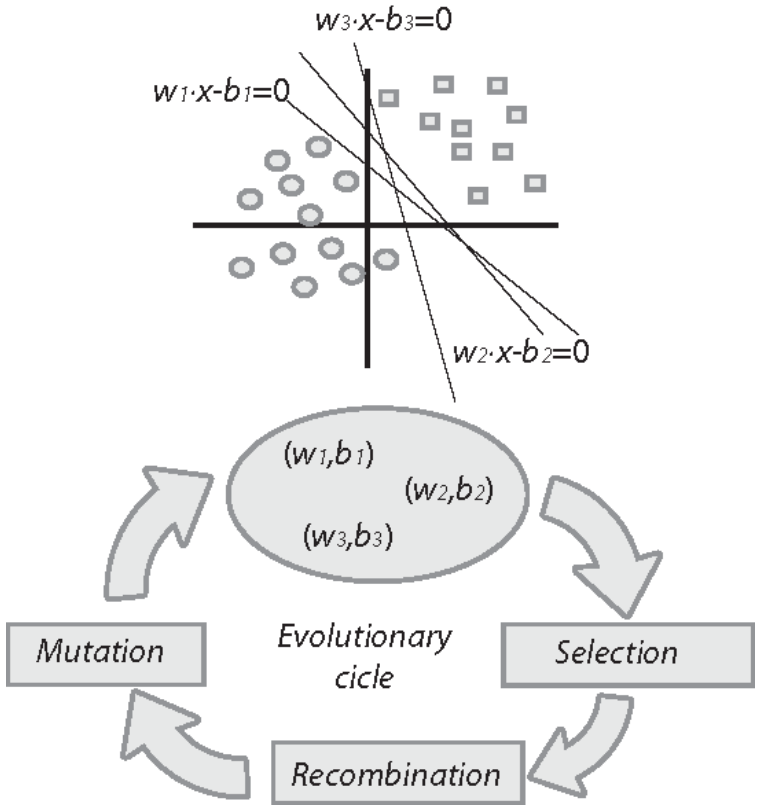
rstoean@inf.ucv.ro

Medicine with machine learning

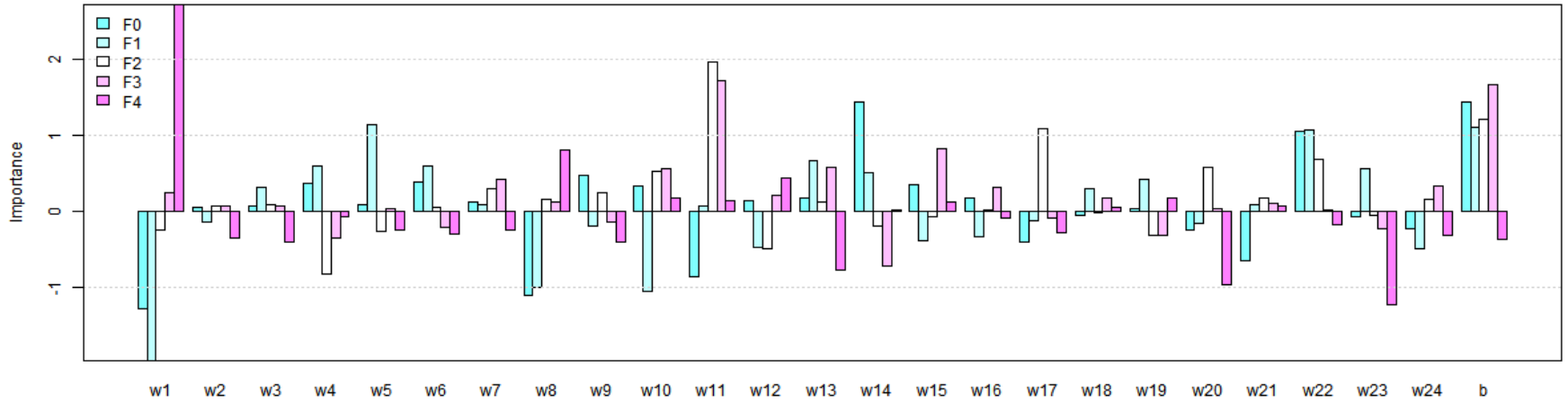
- Different machine learning techniques
  - Support vector machines
  - Neural networks
  - Decision trees
- Enhancement by evolutionary algorithms
  - Optimization
  - Evolutionary rule generation
- Problems:
  - Liver fibrosis stadialization (continuous numerical data)
  - Prediction of length of stay for patients with colorectal cancer (discrete numerical data)

SVM learning, EA optimization

- SVM geometrical separation
- EA individuals encode SVM parameters w and b
- Evaluation of individuals quantifies most accurate and general separation (following SVM principles)
- Output:
  - Weights for attributes
  - Additional simultaneous feature selection through a genetic algorithm at a higher level



Liver fibrosis stadialization

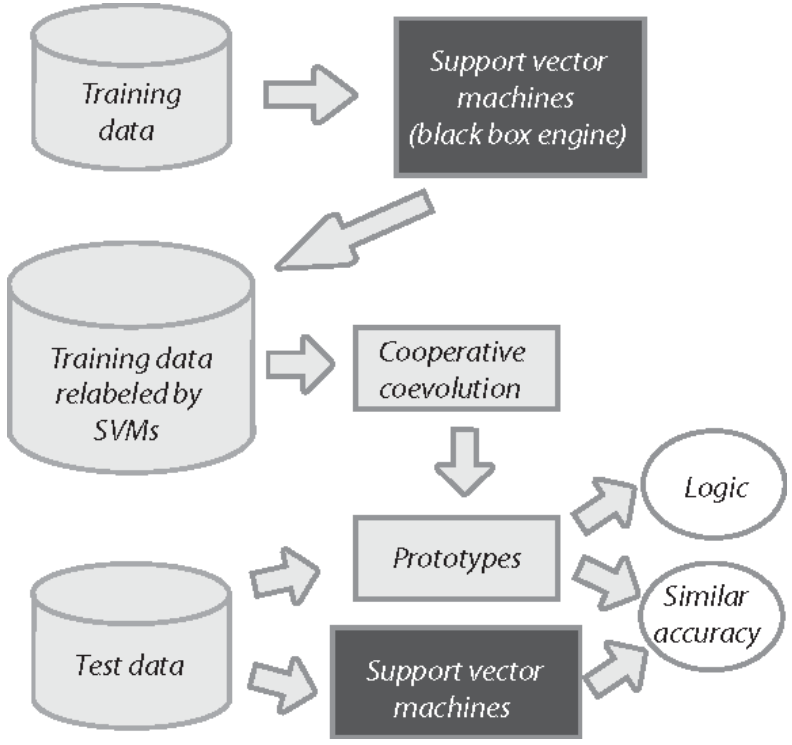


- Available from UMF Cluj
- 722 samples
- 24 **continuous** indicators (stiffness indicator from Fibroscan & hematological and biochemical exams to test the hepatic function)
- 5 grades of fibrosis: normal F0 (29), F1 (227), F2 (164), F3 (87), F4 (215)

Ruxandra Stoean, Catalin Stoean, Monica Lupsor, Horia Stefanescu, Radu Badea, *Evolutionary-Driven Support Vector Machines for Determining the Degree of Liver Fibrosis in Chronic Hepatitis C*, **Artificial Intelligence in Medicine** ,Vol. 51, No. 1, pp. 53-65, 2011.

SVM learning, EA interpretation

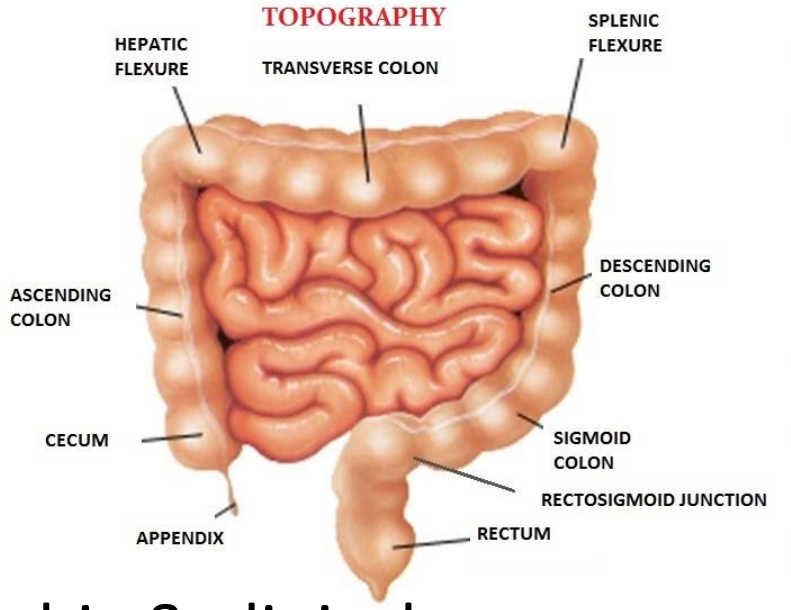
- SVM classification:
  - Pedagogical (changed labels)
  - Decompositional (just support vectors)
- Evolution of prototypes with threshold values through cooperative coevolution (CC)
- Fitness of a class prototype taken with the rules of complementary species as the training set accuracy
- Online and post-feature selection



Ruxandra Stoean, Catalin Stoean, *Modeling Medical Decision Making by Support Vector Machines, Explaining by Rules of Evolutionary Algorithms with Feature Selection*, **Expert Systems with Applications**, Vol. 40, No. 7, pp. 2677-2686, 2013.

Catalin Stoean, Ruxandra Stoean, *Post-evolution of variable-length class prototypes to unlock decision making within support vector machines*, **Applied Soft Computing**, Vol. 25, pp. 159-173, 2014.

LOS in colorectal cancer



Adapted from http://www.medassoc.com/Services/North-Georgia-Endoscopy-Center/Colonoscopy.aspx

- Available from UMF Craiova
- 621 samples
- 8 **discrete** protocol attributes (socio-demographic & clinical – surgery type, tumor parameters, stage)
- 3 classes: short (297), medium (161) and long stay (163)
- Data available at https://doi.org/10.6084/m9.figshare.4747243.v2

Machine learning for LOS

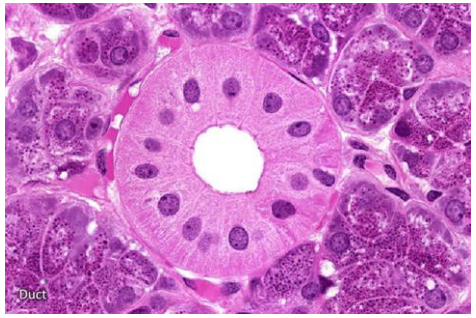
- Support vector machines
- Feedforward neural networks
- C5.0 decision trees
- Ensemble methods: bagging, boosting and random forests
- Evolutionary rule generation
  - An individual encodes all rules (one for each class)
  - Each gene of every rule targets an attribute with a disjunction of possible values for it
  - The class is given by the position in the individual
  - Fitness is the rule set accuracy on the training samples

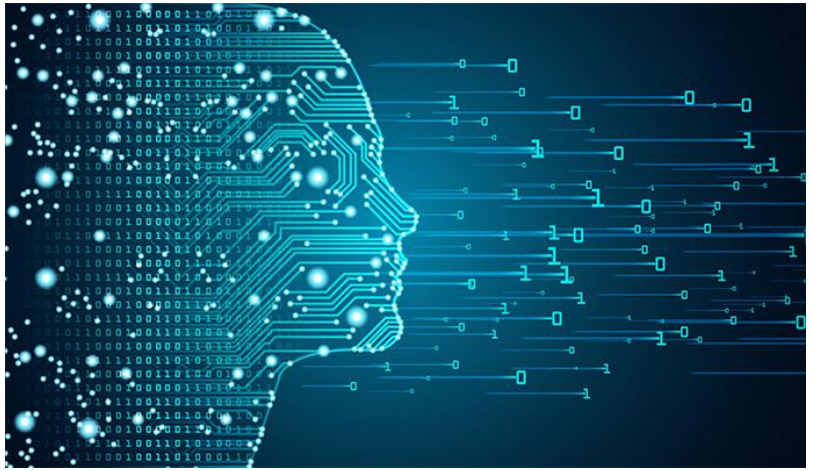
Example rules

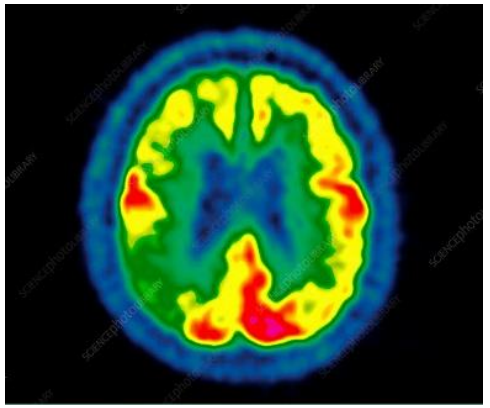
Rule No.	Age	Sex	Topography	Stage	T	N	M	Surgery	Class
DT Rule Set Example									
1	-	-	1	-	-	4	-	256	short
2	-	-	-	-	-	-	-	2345679	short
3	-	-	1	-	34	4	-	379	medium
4	-	-	49	-	-	-	-	1	medium
5	-	-	4	-	34	4	-	25	medium
6	-	-	2789	-	34	4	-	234579	medium
7	-	1	-	-	34	4	-	2345679	medium
8	-	-	167	-	-	-	-	18	long
9	-	-	-	-	-	-	3	2	long
10	-	2	5	-	-	4	-	-	long
11	-	-	259	-	12	-	-	-	long
12	-	-	3	-	-	-	-	37	long
default						short			
EA Rule Set Example									
1	-	-	12678	1	12	3	4	56	short
2	-	-	4	3	3	-	1234	69	medium
3	345	-	15679	34	123	3	-	18	long

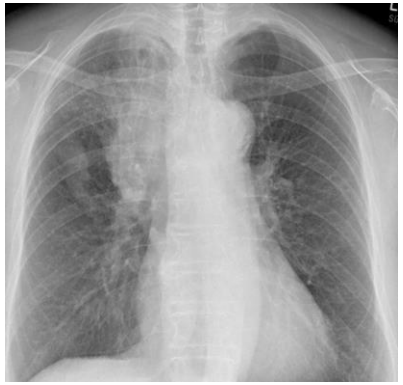
Ruxandra Stoean, Catalin Stoean, Adrian Sandita, Daniela Ciobanu, Cristian Mesina, *Interpreting Decision Support from Multiple Classifiers for Predicting Length of Stay in Patients with Colorectal Carcinoma*, **Neural Processing Letters**, Vol. 46, No. 3, pp. 811-827, 2017.

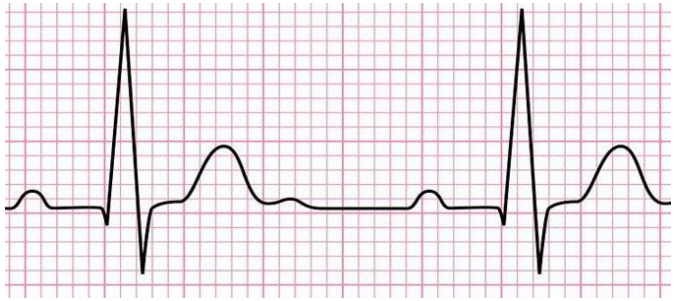
Medicine with deep learning

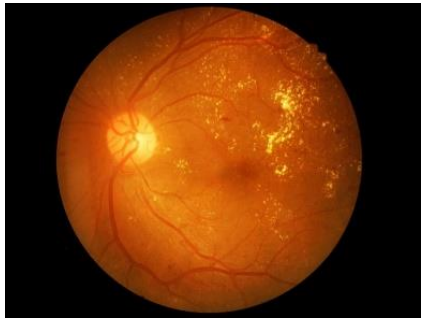










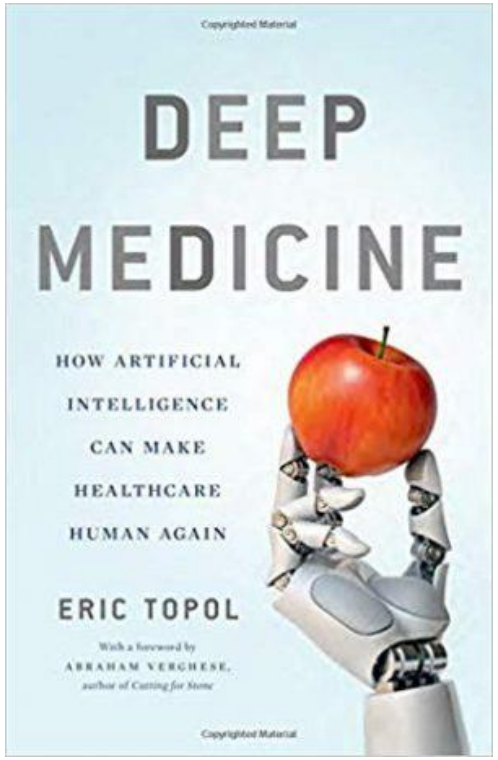


Deep medicine (March 2019)

- Will improve disease diagnosis accuracy
- Virtual medical assistants to clinicians

- But lack of transparency regarding feature extraction
- And susceptibility to adversarial examples

- Ultimately, it could provide more time for doctor-patient interaction



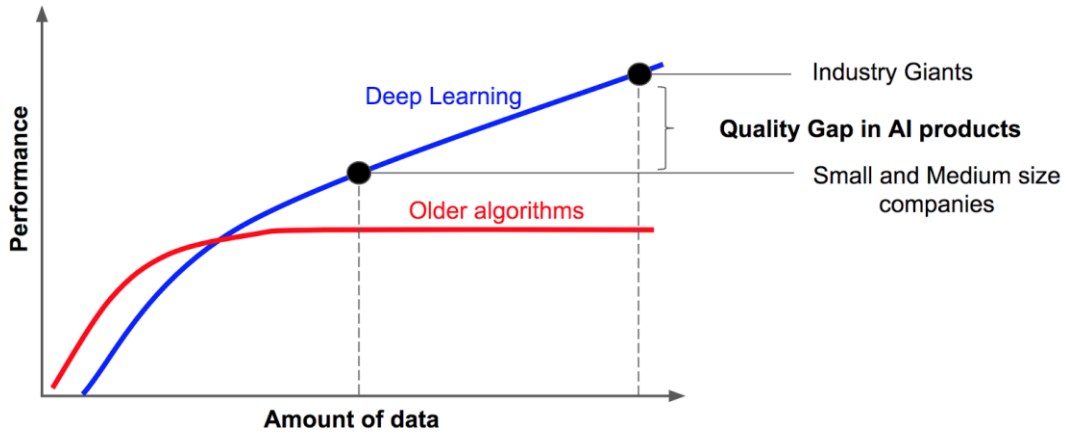
Source: Book and media reviews from the journal Science

Advantages (from own experimentation)

- High prediction power
  - Classification – disease grading
- No requirement of feature indication from physician
- Easy architecture construction
- Straightforward possibility of using transfer learning

Concerns from own experimentation: Data

- *"For most flavors of the old generations of learning algorithms … performance will plateau. … deep learning … is the first class of algorithms … that is scalable. … performance just keeps getting better as you feed them more data"* - Andrew Ng



What people think AI is about

The reality

Concern: Overfitting

- Small available sample size for deep learning application
- When the model is too complex for the data
- Ways to combat:
  - Data augmentation
    - Flip, rotate, scale, crop, translate, Gaussian noise
    - Generative adversarial networks (GAN)
  - Dropout layer
  - Regularization: weight decay - L1 and L2 penalty
  - Early stopping
  - Transfer learning

Interviewer: What's your biggest strength?

Me: I'm an expert in machine learning.

Inteviewer: What's 6 + 10?

Me: Zero.

Interviewer: Nowhere near, it's 16.

Me: It's 16.

Interviewer: Ok... What's 10 + 20?

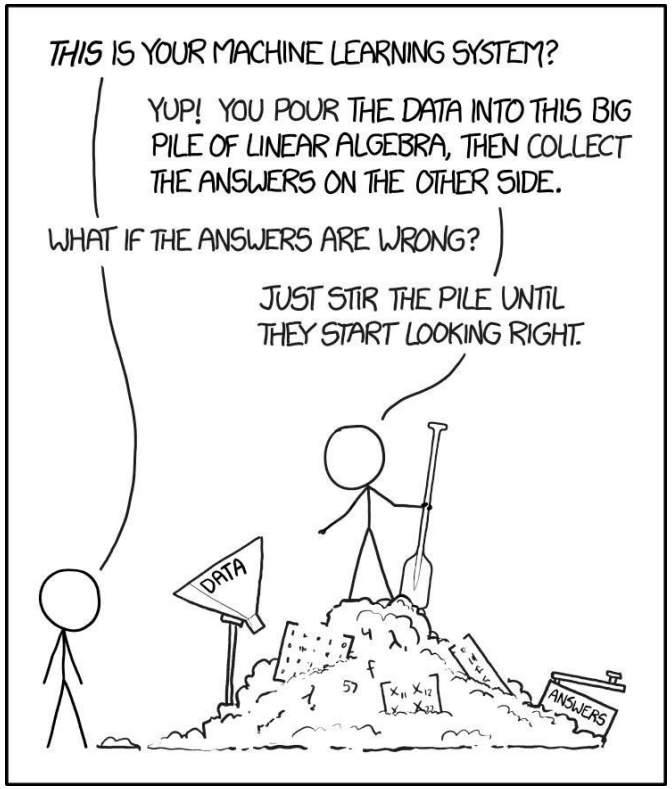
Me: It's 16.

Concern: Data quality

- Few samples for healthy class
- Presence of noise in the data
- Different levels of experience of the operator of the investigation machine
  - Lack of consistency in the set of images/movies
- Different angles when the object of interest is moving – e.g. fetal ultrasound
- Lack of attention when writing down the label for the record
- Taking several samples of the same patient for control:
  - One or two show some presence of illness but all are labelled with the same class
- All these confuse the deep learning model

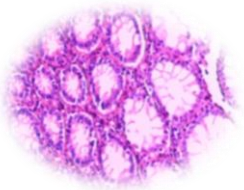
Concern: Parametrization

- Problem-dependent
- Time prohibitive
- Great computing power required
- Automatic (by means of an heuristic)? – TBC

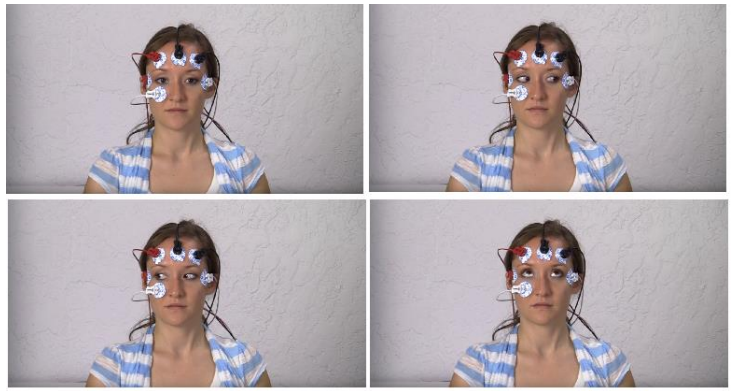


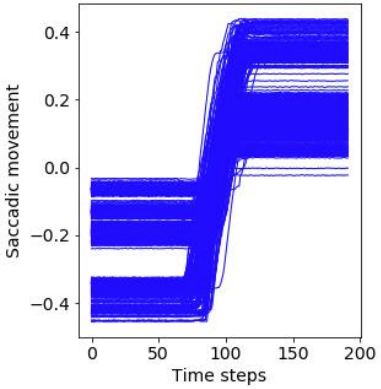
Examples of medical problems for deep modelling

- Colorectal cancer grading from histopathological slides
  - Images



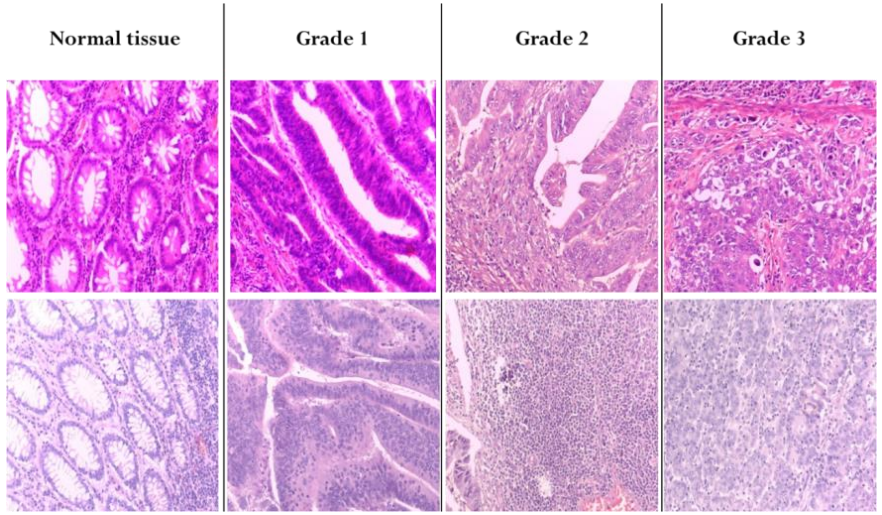
- Grading of spino-cerebral ataxia of type 2 (SCA2)
  - Time series

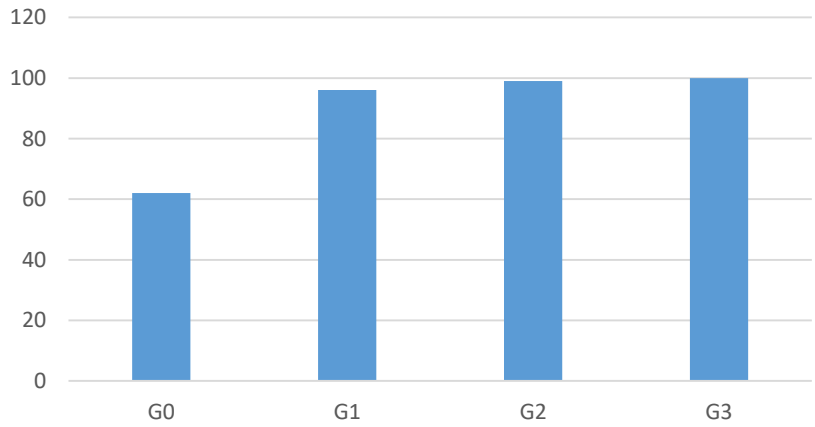




Colorectal cancer grading

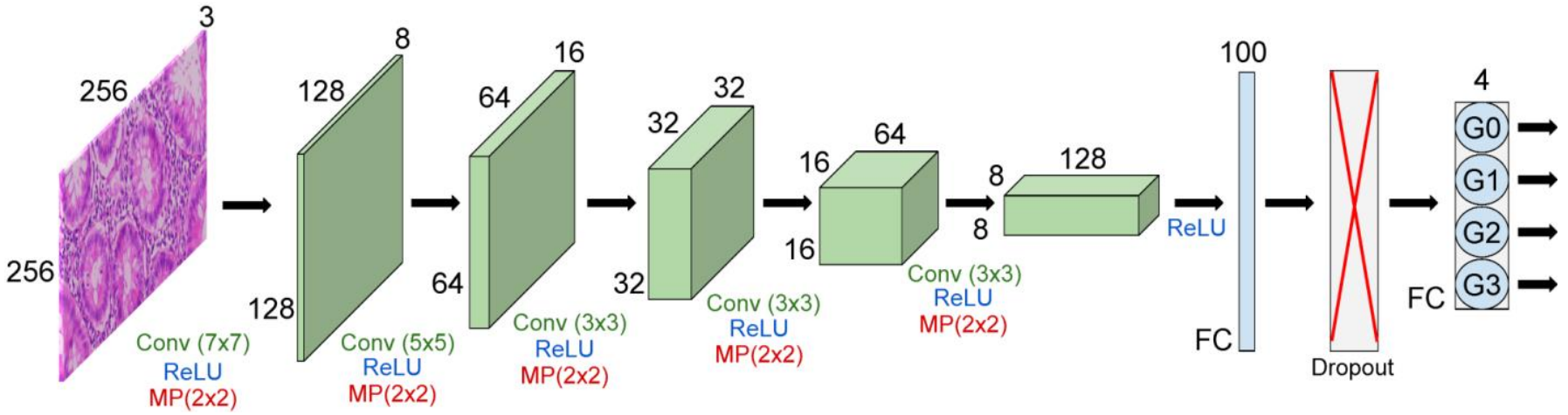
- 357 samples, **images** 800x600 pixels
- 4 cancer grades: G0 (normal), G1-G3
- Available from UMF Craiova
- https://doi.org/10.6084/m9.figshare.4508672.v1





CNN architecture

- Python with Tensorflow framework
- Input: volume of 256x256x3 (width x height x color channels)
- Output: 4-dim. vector holding scores for each class
- Convolution -> ReLU -> Max-Pooling (5 cycles)
- Fully-connected layer -> Dropout layer -> Fully-connected layer



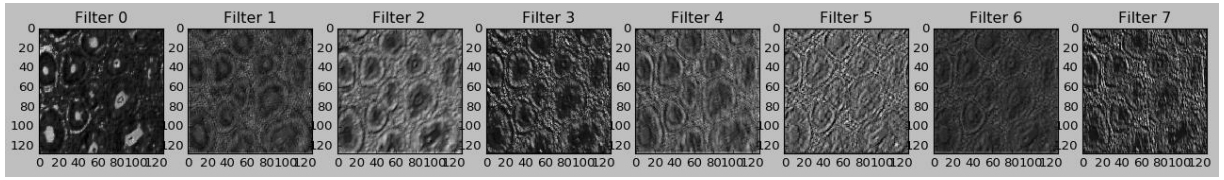
Results

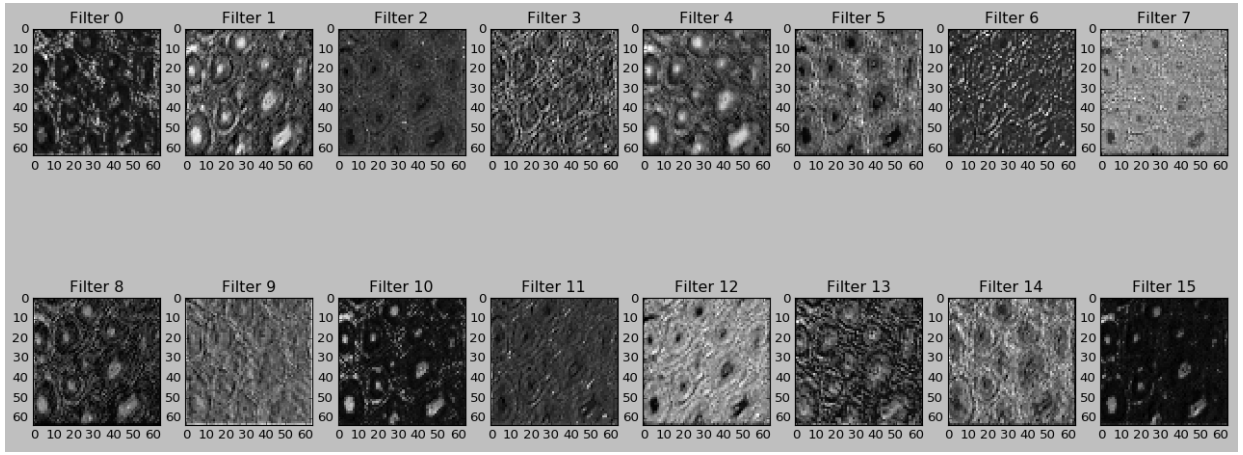
- Compared to multi-stage approach:
  - Segmentation of glands and nuclei
  - Feature extraction leading to 76 numerical features
  - SVM accuracy 79.89%
  - SVM + GA feature selection 83.94%

Catalin Stoean, Ruxandra Stoean, Adrian Sandita, Cristian Mesina, Corina Lavinia Gruia, Daniela Ciobanu, *How Much and Where to Use Manual Guidance in the Computational Detection of Contours for Histopathological Images?*, **Soft Computing**, Vol. 23, No. 11, pp. 3707–3722, 2019.

- CNN - 90.15%
  - G1 and G3 with no error
  - 9% from G0 labelled as G2
  - G2 with an accuracy of only 73% with the majority labeled as G1 and none normal

Activation maps for a slice with normal tissue



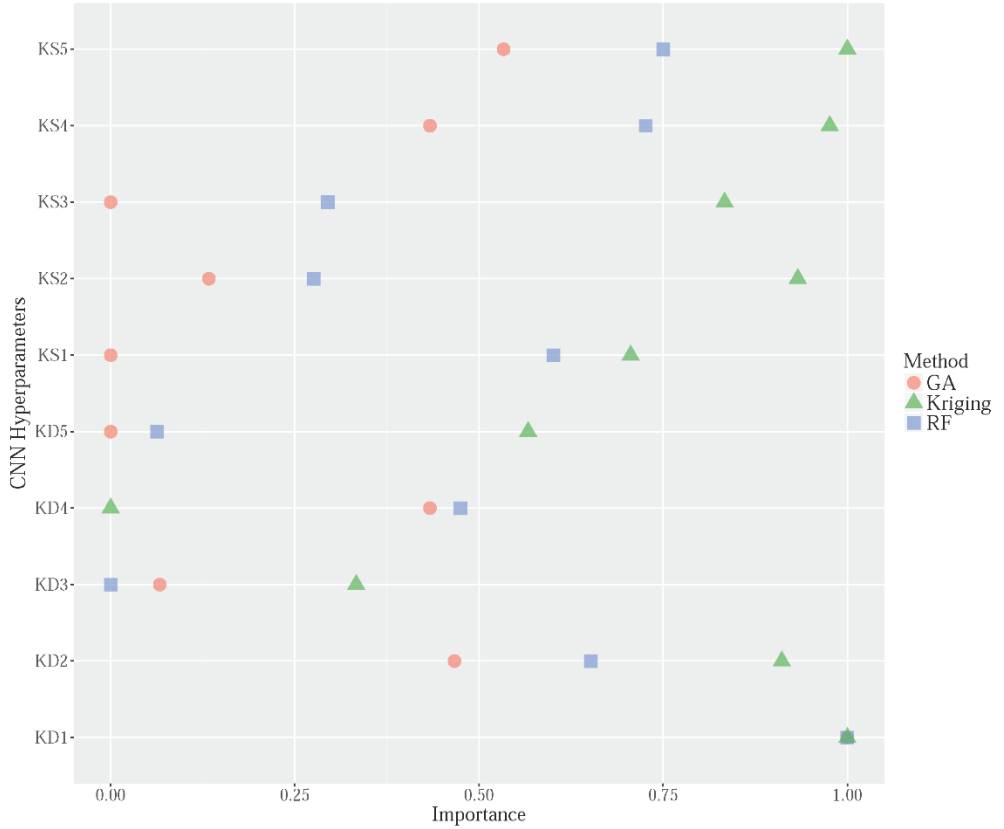


Surrogate-EA offline CNN parametrization

- Latin hypercube sampling
  - Generate a set of diverse configurations (100) for kernel sizes and depths
  - Record for each setting the CNN validation accuracy from 10 runs with random split of data
- Surrogate regression models learn the pattern from the generated data
  - Support vector machines, random forests, Kriging
- Evolutionary algorithms
  - Search for new configurations
  - Fitness given by estimation of surrogates
  - Update with real CNN prediction of best individual every few generations
- CNN with parameters of best EA individual and a SVM surrogate gives a test accuracy from 30 runs of 92%

Variable importance

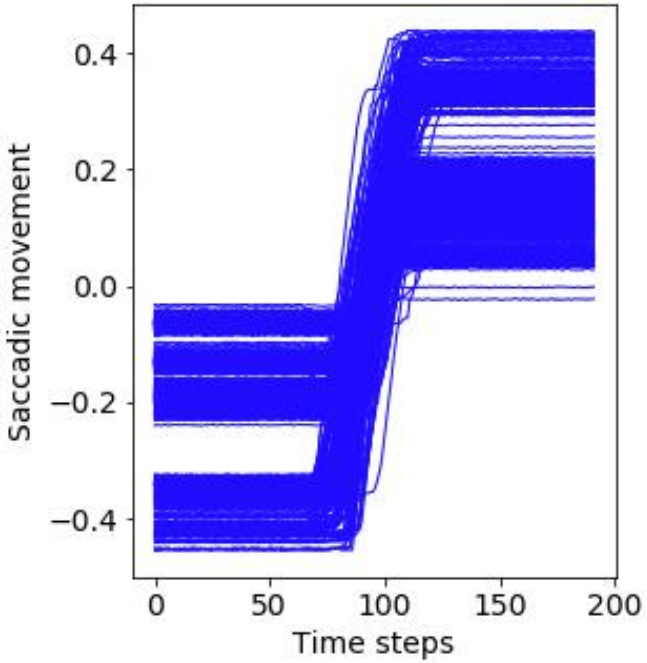
- LHS data with CNN variables and accuracy
- A genetic algorithm - individuals encode the presence/ absence of every variable
- Fitness of selected attributes = SVM error estimated from k-fold cross-validation on the training LHS data
- The GA is repeated a number of times
- The score for each CNN variable is the number of times selected in the best individual



Ruxandra Stoean, *Analysis on the potential of an EA–surrogate modelling tandem for deep learning parametrization: an example for cancer classification from medical images*, **Neural Computing and Applications**, 2019, https://doi.org/10.1007/s00521-018-3709-5

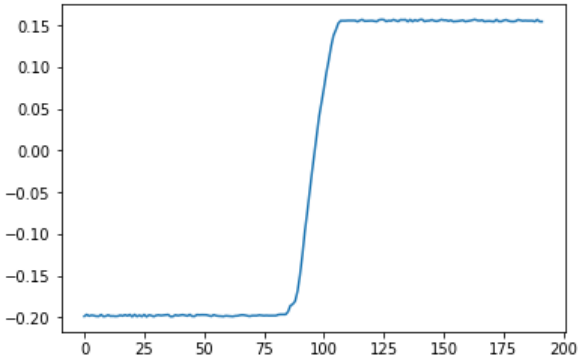
Saccadic eye movement

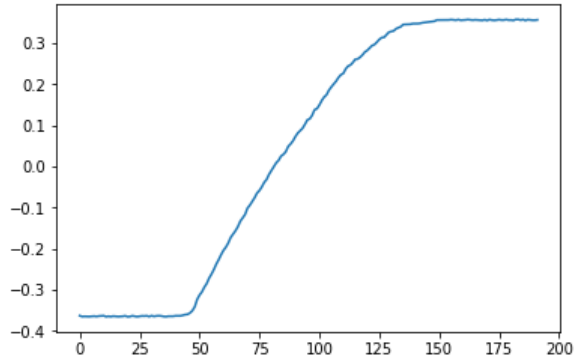
- Spino-cerebellar ataxia of type 2 (SCA2)
  - affects the nervous system
  - can be diagnosed by genetic analysis – expensive
  - electrooculography: examining the weak electrical potentials generated by the eye movement (saccades) of a person tracking the trajectory of an object

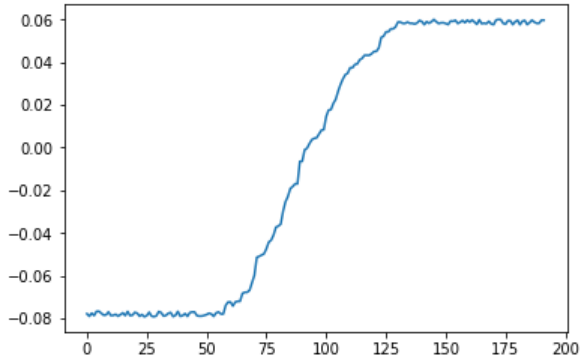


Classification

- Healthy, presymptomatic and ill cases
- Early detection of presymptomatic cases triggers a timely medical intervention
- Provided by the Center for the Investigation and Rehabilitation of Hereditary Ataxias, Cuba
- In cooperation with the University of Malaga

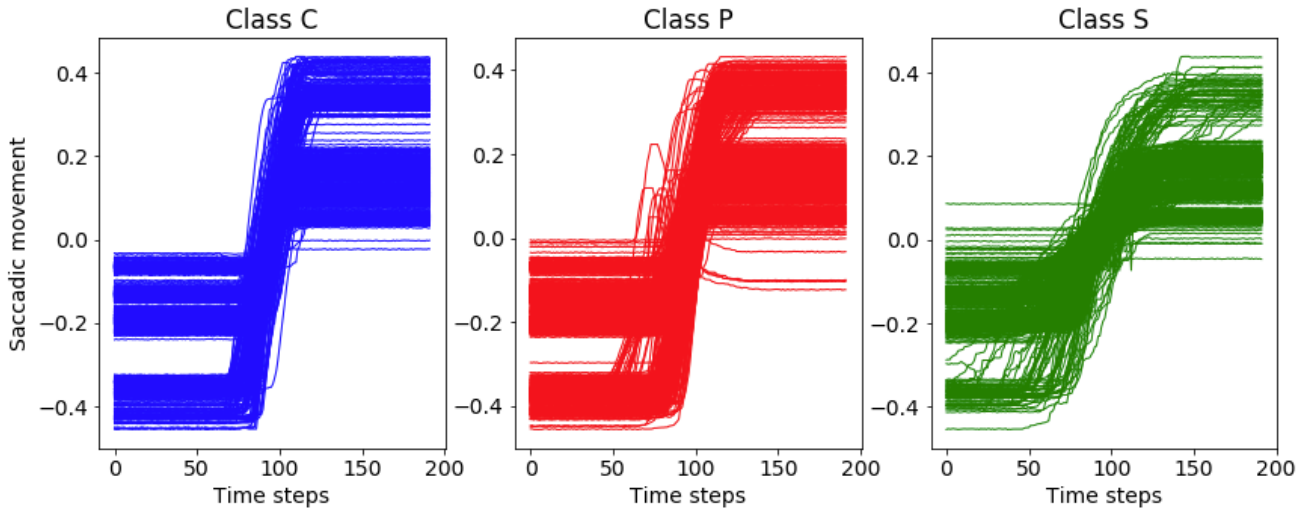






Saccades and registers

- An electrooculographic test consists of several evaluations
- More saccadic samples form a register, which is labeled by the expert
- The majority of saccades in the register of an ill person are clearly distinct from those of a healthy person
- Most saccades from a presymptomatic register resemble those of a healthy person
  - But some have a slightly different form



Data set

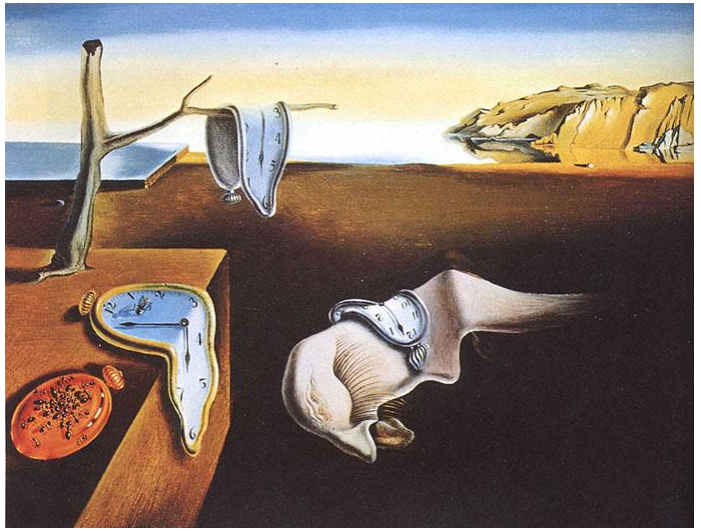
- Registers from 88 subjects:
  - 29 C (healthy, control)
  - 29 P (presymptomatic cases)
  - 30 S (ill, sick)
- 6124 saccade arrays of length 192
- Split into training-validation-test as registers (and saccades):
  - C: 12-12-5 registers (saccades 901-977-360)
  - P: 12-12-5 (1039-845-421)
  - S: 12-12-6 (491-771-319)

Models for saccade recognition

- Supervised learning
  - 1D Convolutional Neural Networks (CNN)
- This type of learning may be confused in establishing the correspondence between the saccades and the register class
- Hence, also unsupervised learning
  - K-means (KM)
  - Self-organizing maps (SOM)
- Always referring to registers and not only individual saccades

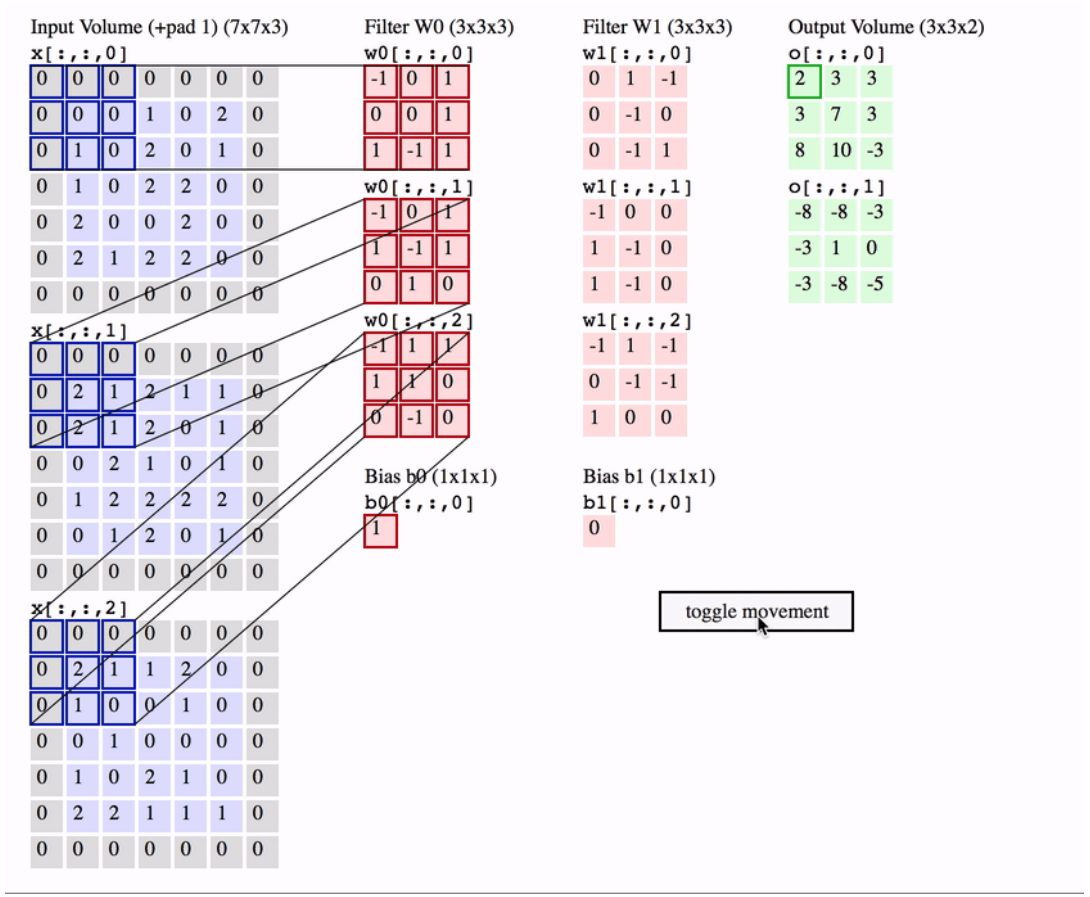
CNN

- LSTM initially tried but too slow
- 1D CNN capable of persistence of memory, faster and similar performance
- Architecture
  - Convolution (KS = 3, KD = 128): 2 cycles
  - Dropout
  - MaxPooling (size = 2, strides = 2)
  - Convolution (KS = 3, KD = 256): 2 cycles
  - Dropout
  - Dense
- Test saccades are labelled by the CNN
- The majority label in a register gives the class



Convolution

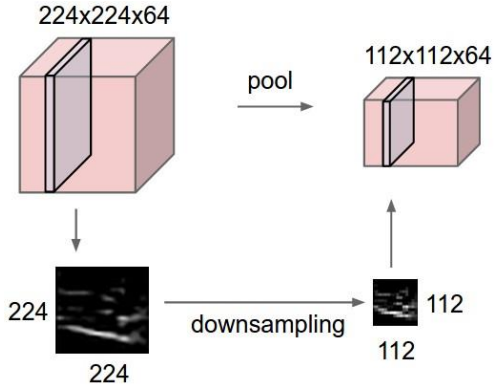
- Between filter and input volume
- Input
  - Volume of Width x Height x Depth
- Kernel (or filter) - a shared set of weights

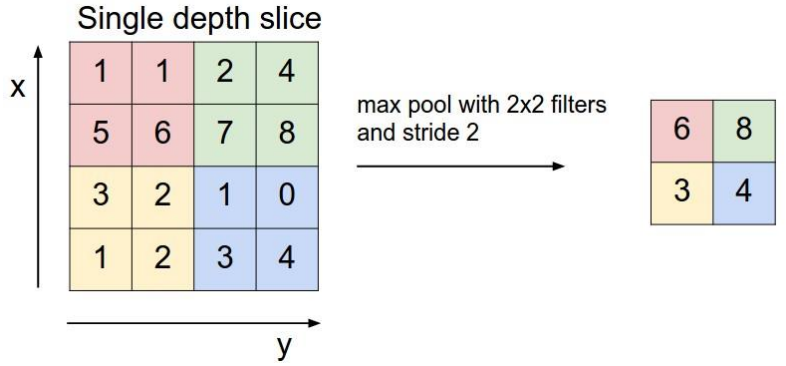


http://cs231n.github.io/convolutional-networks/
https://www.youtube.com/watch?v=AQirPKrAyDg

ReLU & Pooling

- Rectified Linear Unit – transfer layer for nonlinearity
- (Max) Pooling – downsamples the volume
  - Window (pool) size
  - Stride

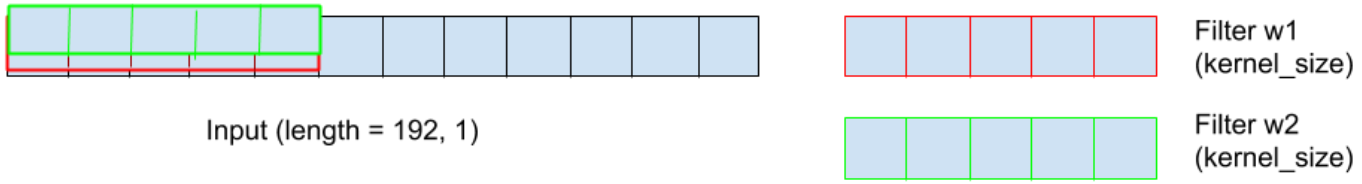




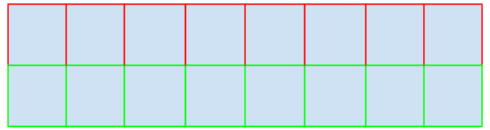
http://cs231n.github.io/convolutional-networks/
https://www.youtube.com/watch?v=AQirPKrAyDg

CNN 1D

- Transformation from 2D to 1D



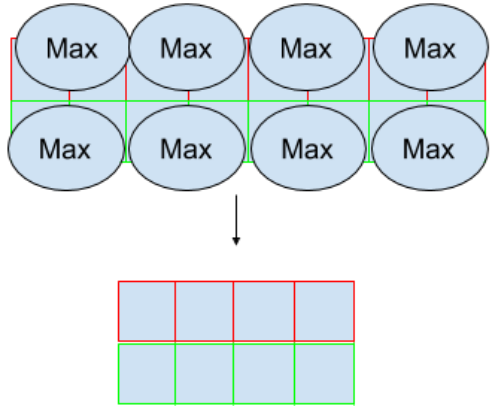
Conv1D



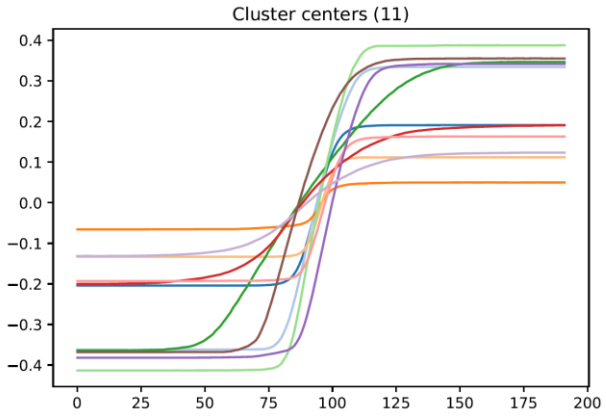
Output (length - (kernel_size - 1), noOfFilters)

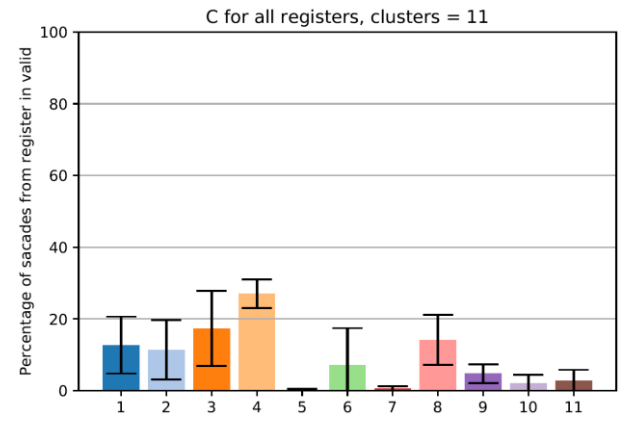
MaxPooling1D

pool_size = 2
stride = 2

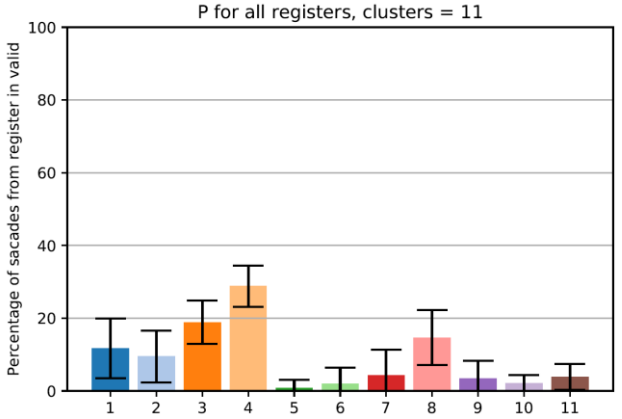


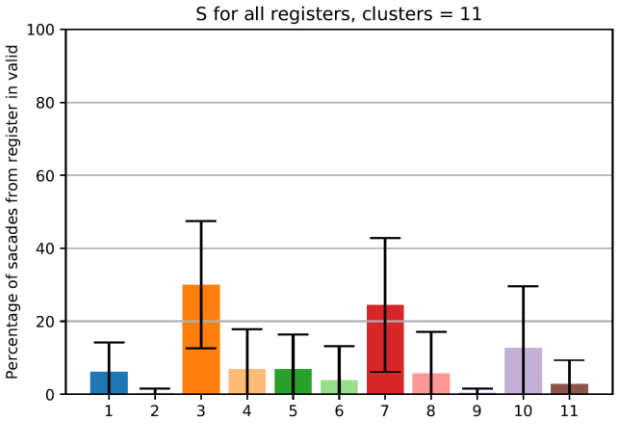
K-Means 1/4





- Search for different profiles of saccades
- Tried various values for k



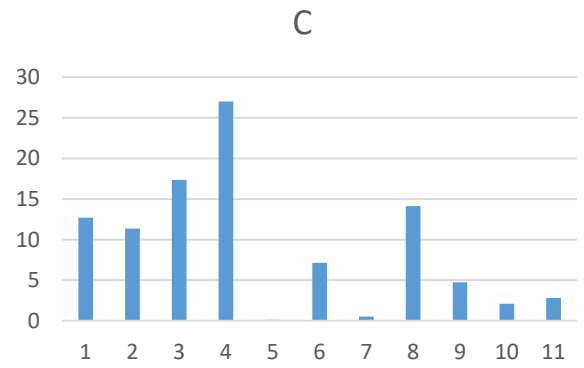


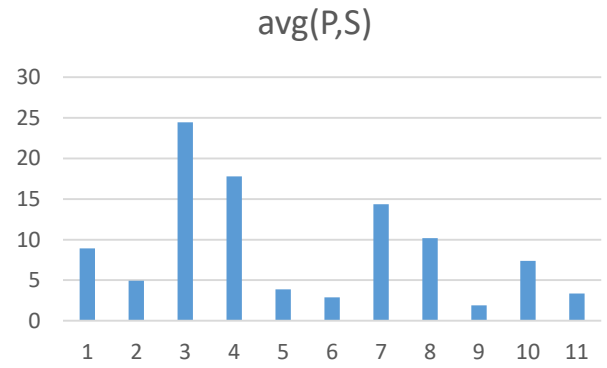
1. Begin by determining the k cluster centroids from the training set, without considering the register class
2. In the validation set, for every label, find the distribution of saccades to each cluster
   - Euclidean distance is used for attribution

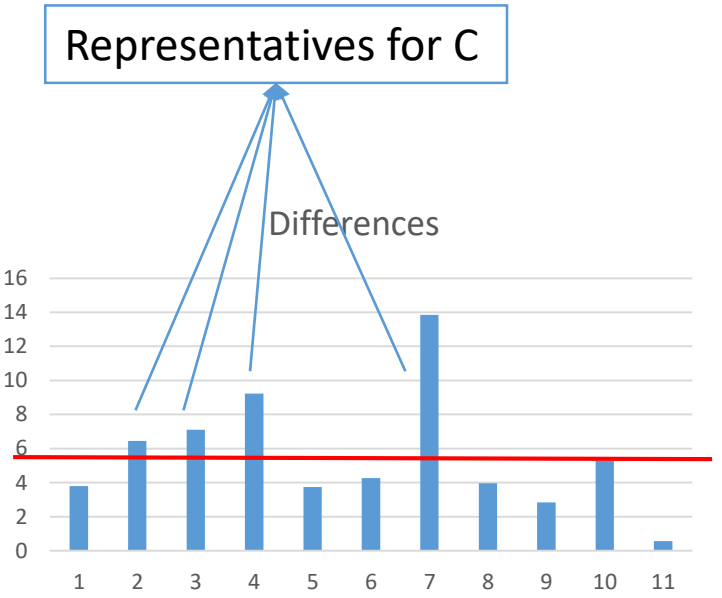
K-Means 2/4

3. The specific characteristics for each class are next identified
   - E.g., more (or less) saccades associated with a cluster
- Example for C:
   a) Compute the average between P and S for each cluster in turn
   b) Compute the absolute difference between C and avg(P, S) for each cluster in turn
   c) Compute average over all clusters for the array obtained in b)

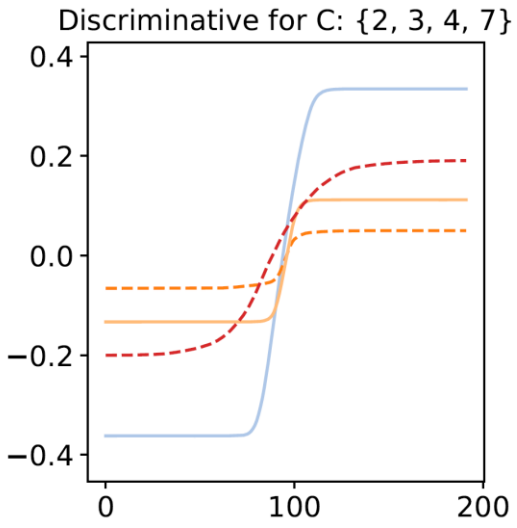
K-Means 3/4 - Example for C

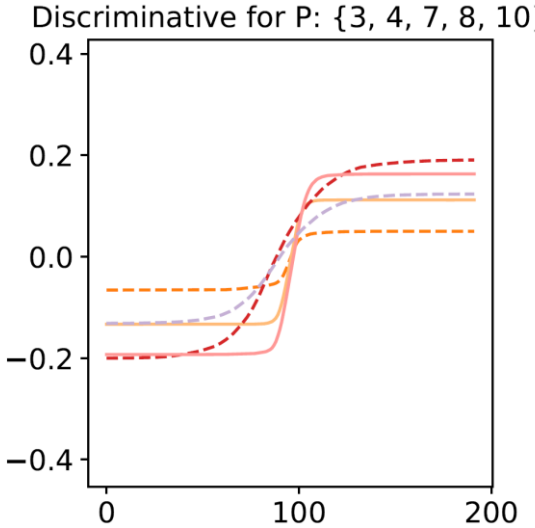


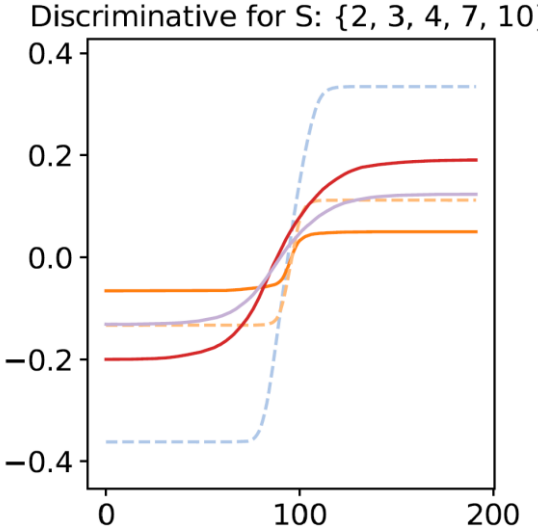




- Similarly, representatives are found for P and S.





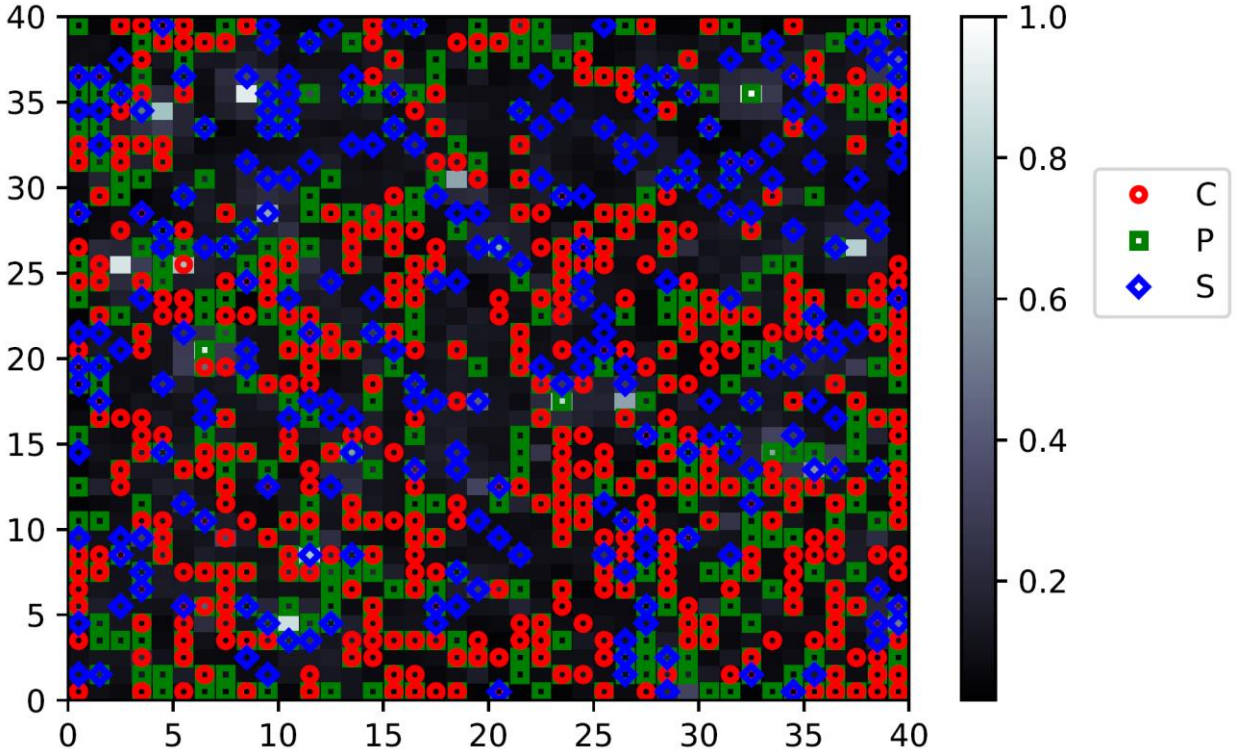


K-Means 4/4

4. The registers from the test set are checked against the discriminative centroids detected in step 3
   - The form of the register (percentage of saccades for each cluster) is obtained
   - Euclidean distance is used between form and profile of each class
   - Distances are divided by the number of centroids found important for each class
   - The label of the test register is given by the class that led to the smallest obtained value

Self Organizing Maps (SOM)

- A different visualization angle than KM
- Generates a two-dimensional map from the initial high-dimensional input of the unlabeled training saccades

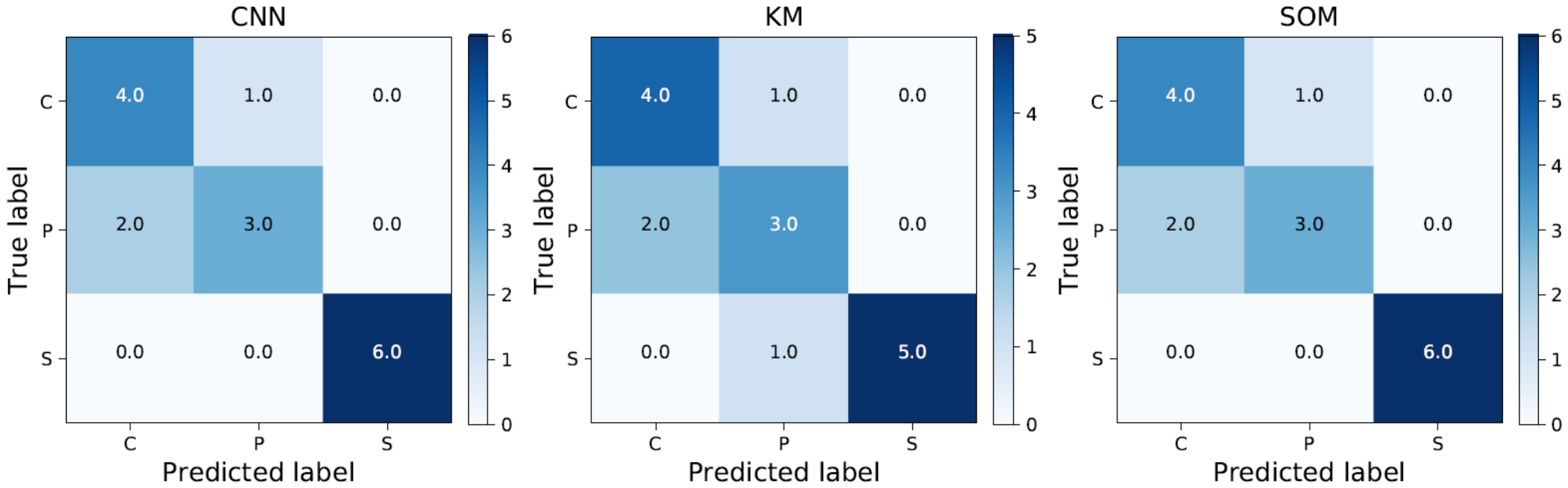


SOM

- The winning neuron is found for each saccade in a test register
- The test saccade thus gets a triplet of votes
  - They correspond to votes for each label of the training samples found in that position
- The test register is assigned the sum of the triplets of each of its saccades
- The class with the maximum value gives its final label
- In case of equality, the leftmost (less severe) condition is taken as the label

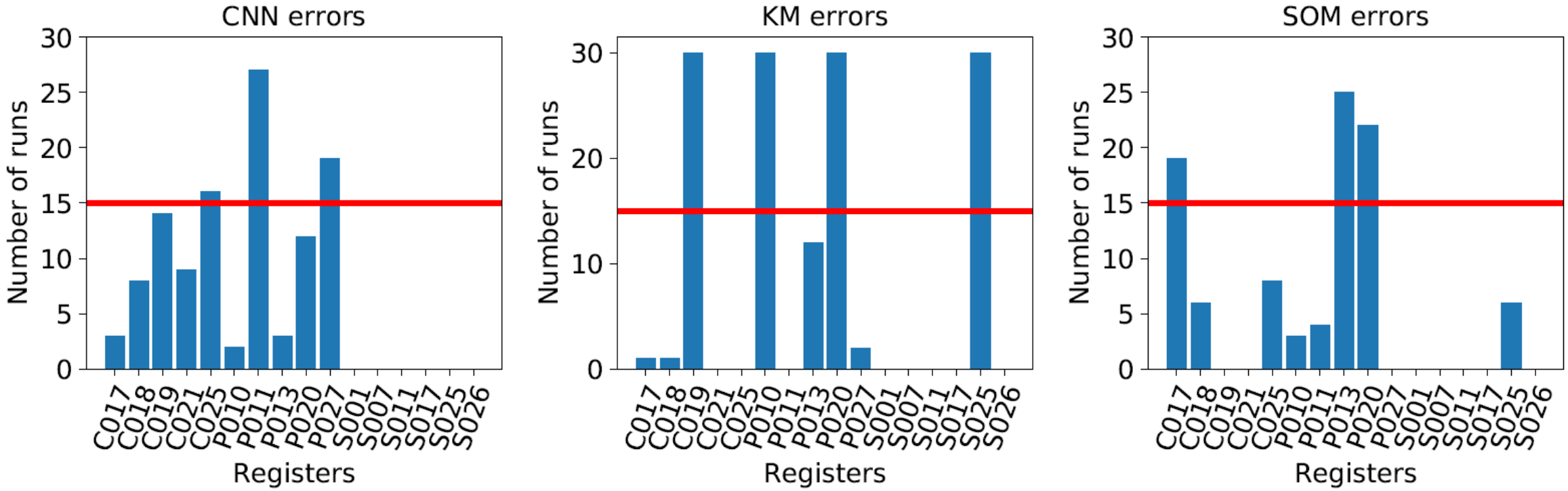
Results 1/2

- Confusion matrices for CNN, KM and SOM
- Each model confuses 2 out of 5 registers for presymptomatic



Results 2/2

- The number of times the class of each test register is mistaken out of 30 different runs.
- Register P020 is mistaken by both KM and SOM

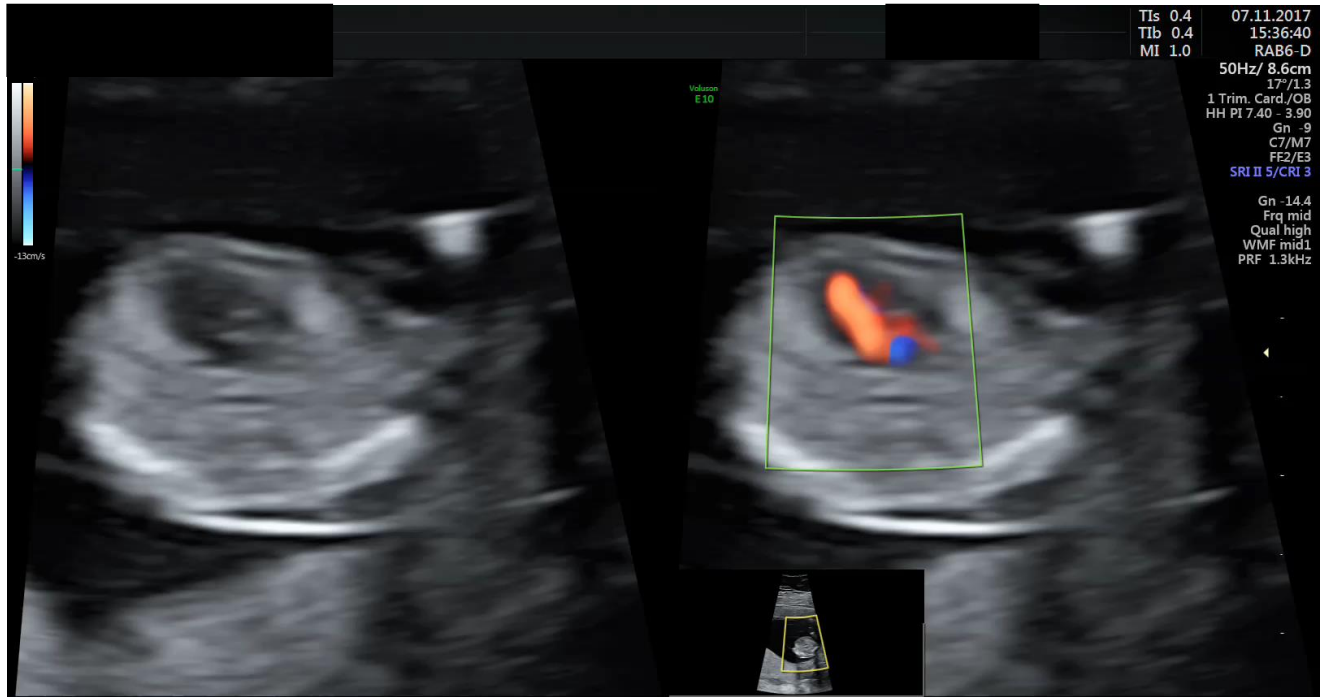


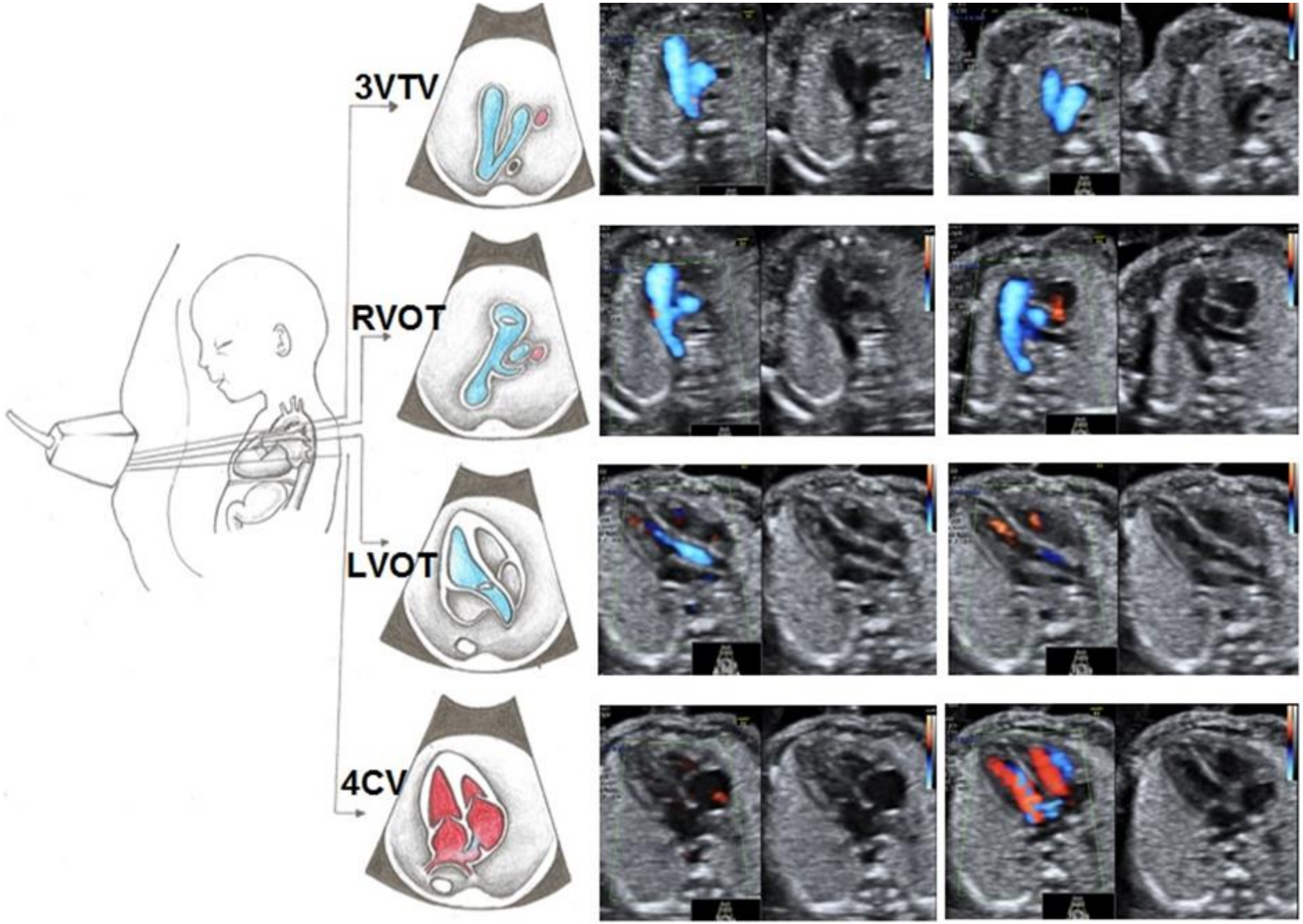
Conclusions

- KM performed worst: 4 registers wrongly classified
- CNN and SOM mistake 3 registers each: 2P and 1C
- Since there is 1 common register for KM and SOM, an ensemble of the 3 would mistake only this register
- Data could be enriched with genetic markers
- Work presented at IWANN – International Work-Conference on Artificial Neural Networks http://iwann.uma.es/

Current work: fetal cardiac analysis

- Instance segmentation
  - There are specific objects in this image and these are the pixels of each




3VTV
RVOT
LVOT
4CV

Deep learning for medicine: future goals

- A better physician <-> computer scientist communication

-> More and better labeled data

"I've always believed that the data is trying to tell us what the answer is, but we need to know how to listen to it. That's where the computation comes in.", Brian Druker, Knight Cancer Institute

<-Explainable models

"I think people need to understand that deep learning is making a lot of things, behind-the-scenes, much better.", Geoffrey Hilton


Q
A