Apathy diagnosis by analyzing **facial dynamics** in videos

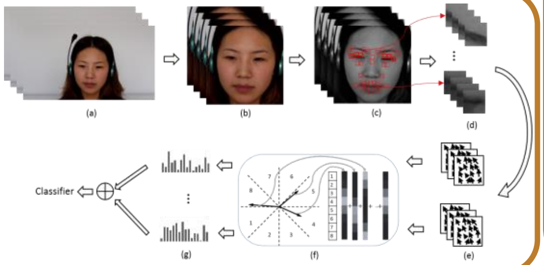
S L Happy

Collaborators: Antitza Dantcheva, Abhijit Das, Radia Zeghari, Philippe Robert, and Francois Bremond

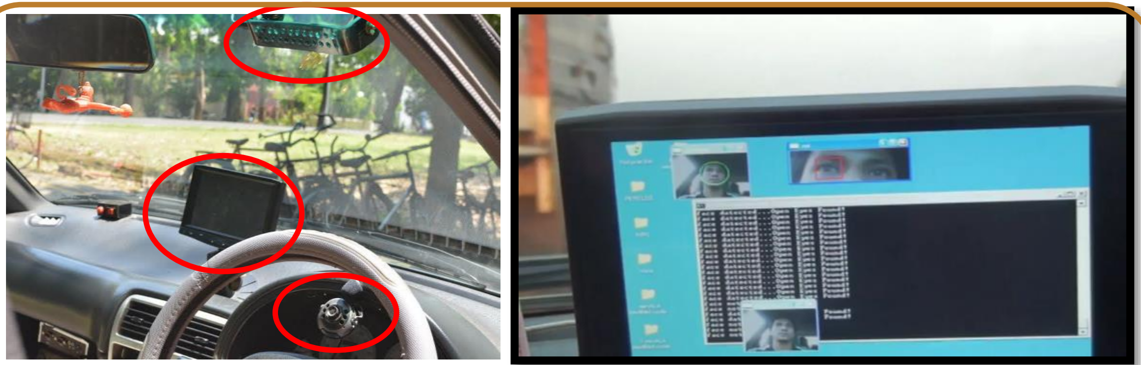
STARS team, **INRIA** Sophia Antipolis

2nd April 2019

My research



(Emotion recognition)



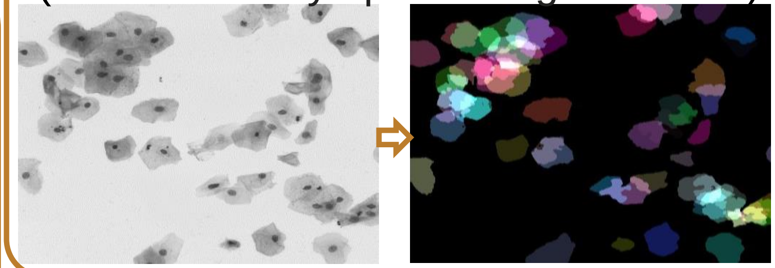
(Driver drowsiness detection system)

Original

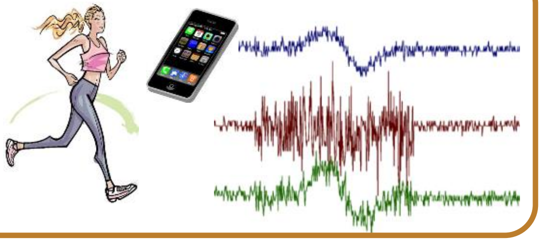
Adapted

(Hyperspectral image classification)

(Cancer Cell cytoplasm segmentation)



(Activity recognition)



Apathy

- Apathy is a symptom of the majority of neurocognitive, neurodegenerative, and psychiatric disorders
- Quantitative reduction of activity in behavioral, cognitive, emotional, or social dimensions
- Symptoms:
  - **reduced emotional response**
  - lack of motivation
  - change of behavior
  - **limited social interaction**

Apathy

Non-Apathy

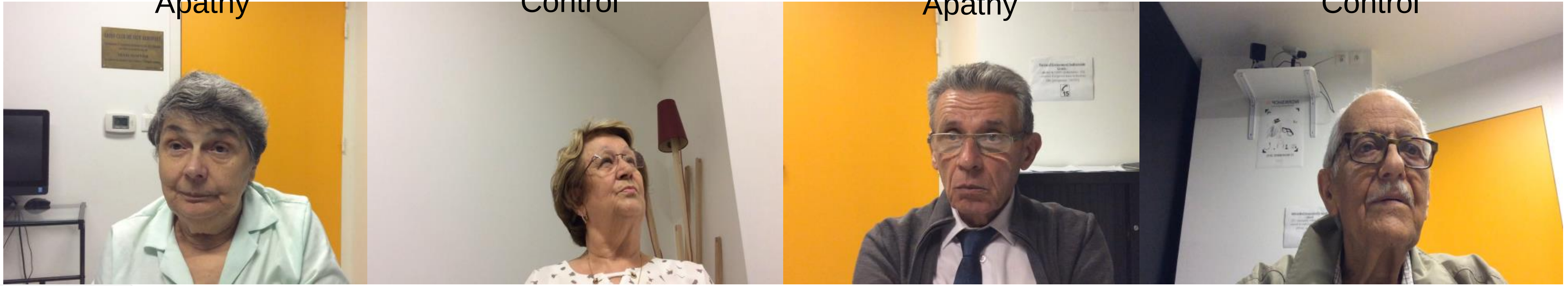
Challenge

Activity classification

Boxing | Pull up | Playing Guitar | Typing

Apathy classification

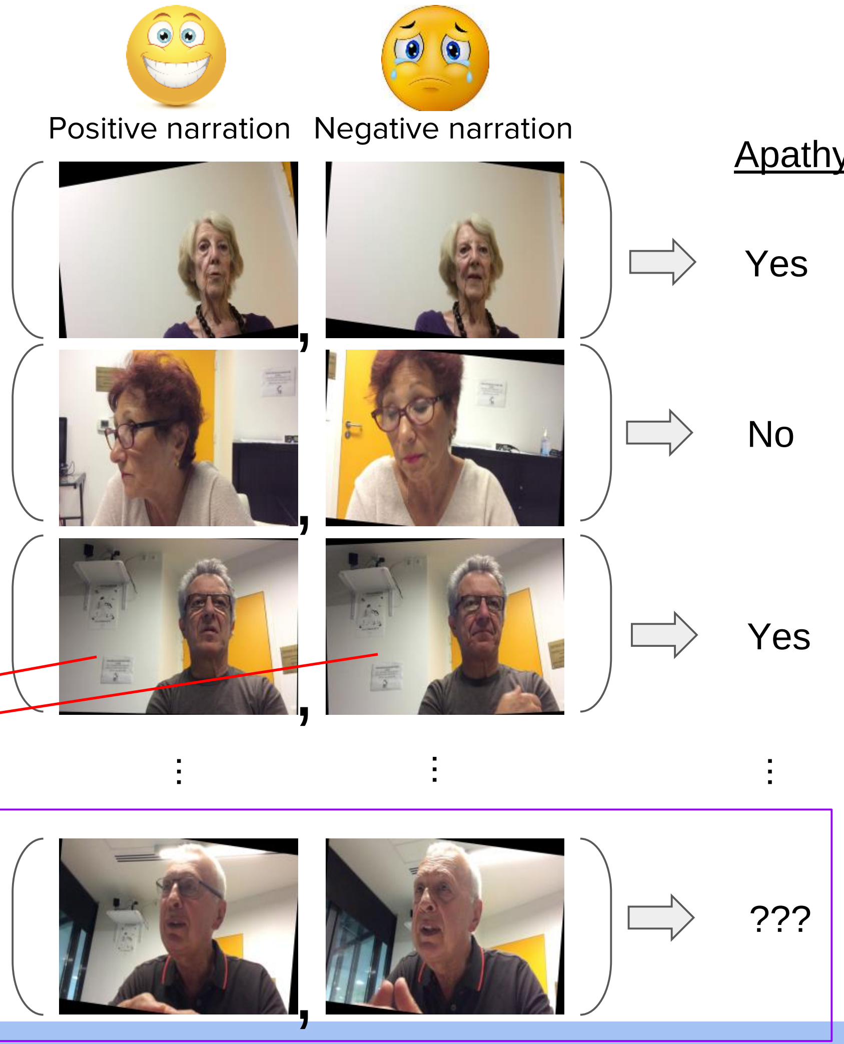
Apathy | Control | Apathy | Control



Objective

- Assisting clinicians in the apathy diagnostics based on facial behavior analysis
- Given the facial videos, predict whether the patient is apathetic or not.

(“tell me a positive/negative event of your life in one minute”)



Part 1: Feature → Regression → Classification

Data

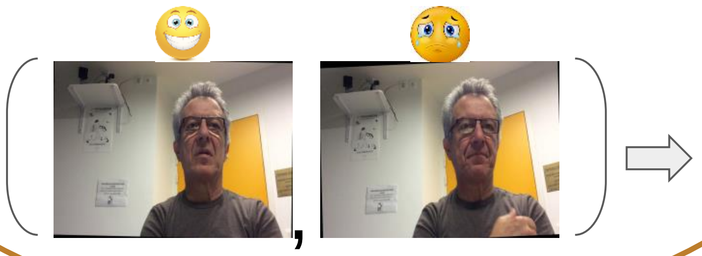
- Total subjects: 45
- The patient-clinician interview involves
  - the collection of demographic details
    - Age, **gender**, ...
  - a standardized neuropsychological assessment
    - Mini mental state examination (**MMSE**)
    - Neuropsychiatric apathy inventory (**NPI-apathy**)
  - a short **positive and negative** experience **narration** (tell me a positive/negative event of your life in one minute)

Demographic data of patients

	Number of Patients	Age	MMSE	NPI-Apathy
Apathy	18	73.5 (7.7)	22.6 (3.1)	6.2 (2.6)
Control	27	71.7 (8.8)	25.4 (3.6)	0.4 (0.8)

Overall Framework

Objective



Training phase

Inputs
- Video of positive narration
- Video of negative narration
- Class labels
- Male/female labels
- Clinical scores

Model training

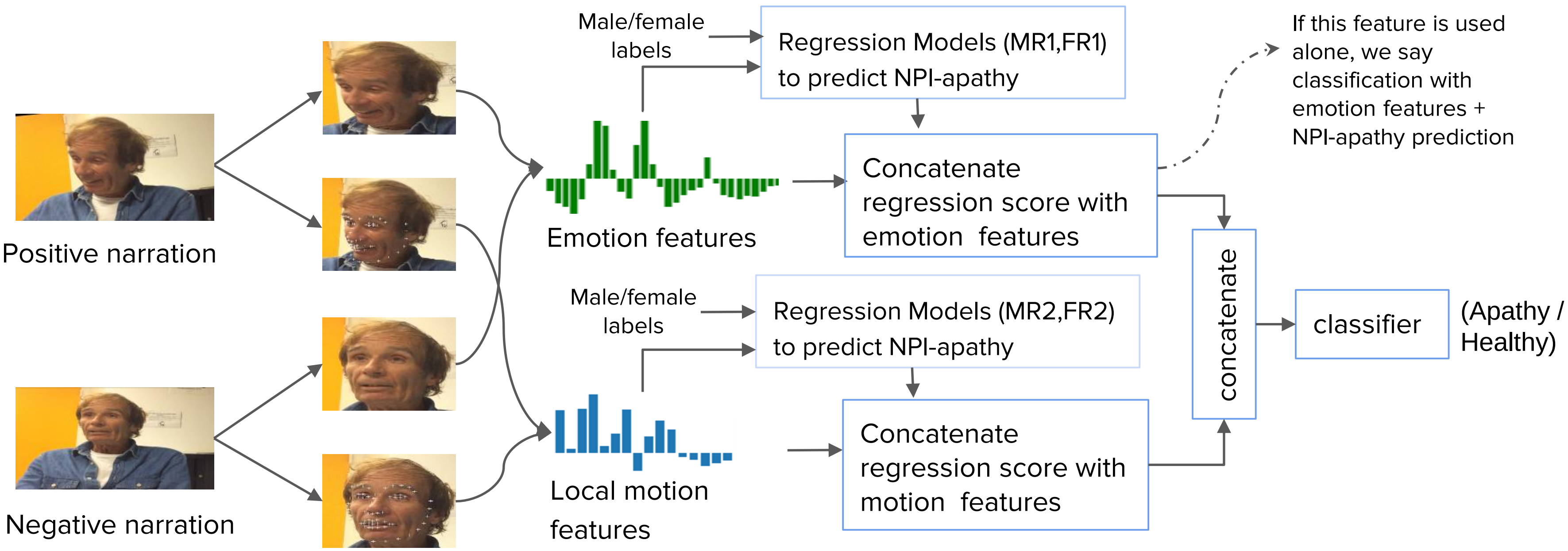
Test phase

- Video of positive narration
- Video of negative narration

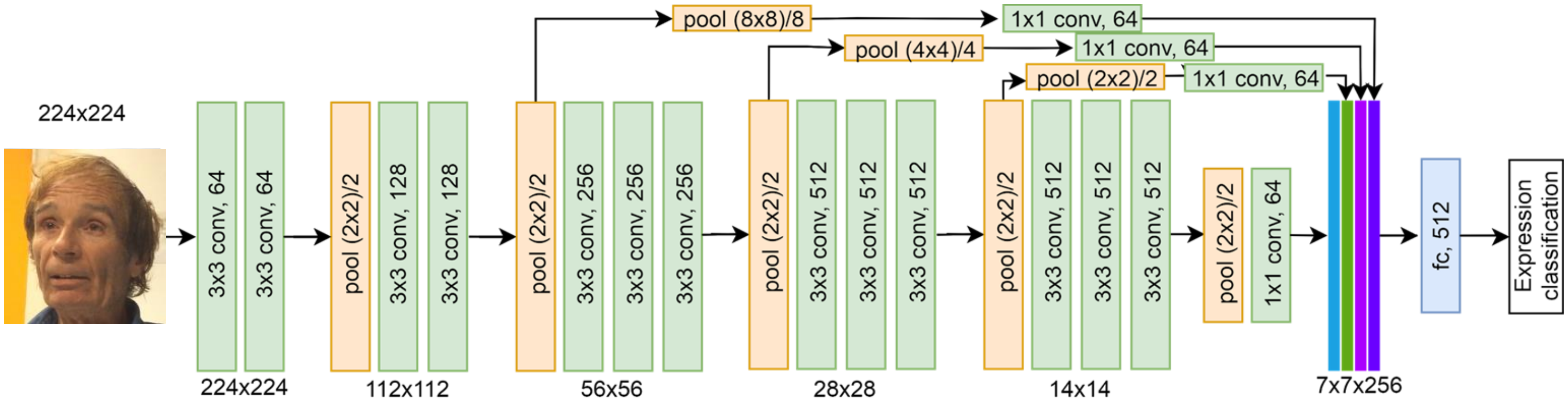
Classification

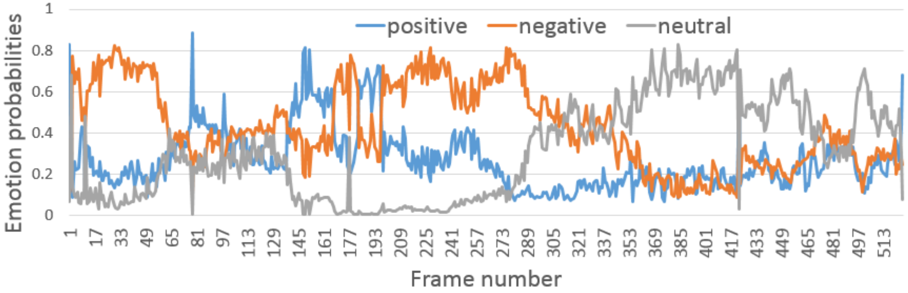
- ❑ Apathy
- ❑ Non-apathy

Detailed Framework



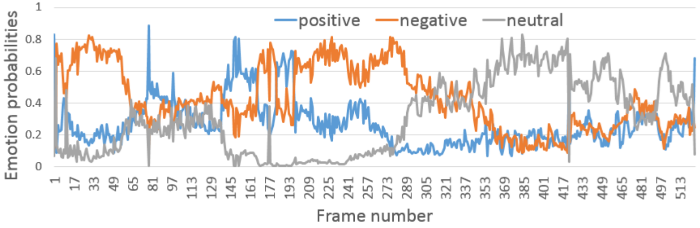
Emotion Recognition



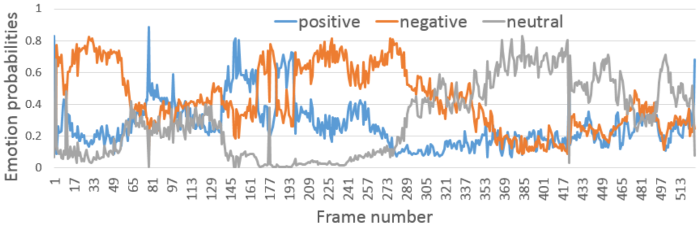


Emotion Feature

Positive narration

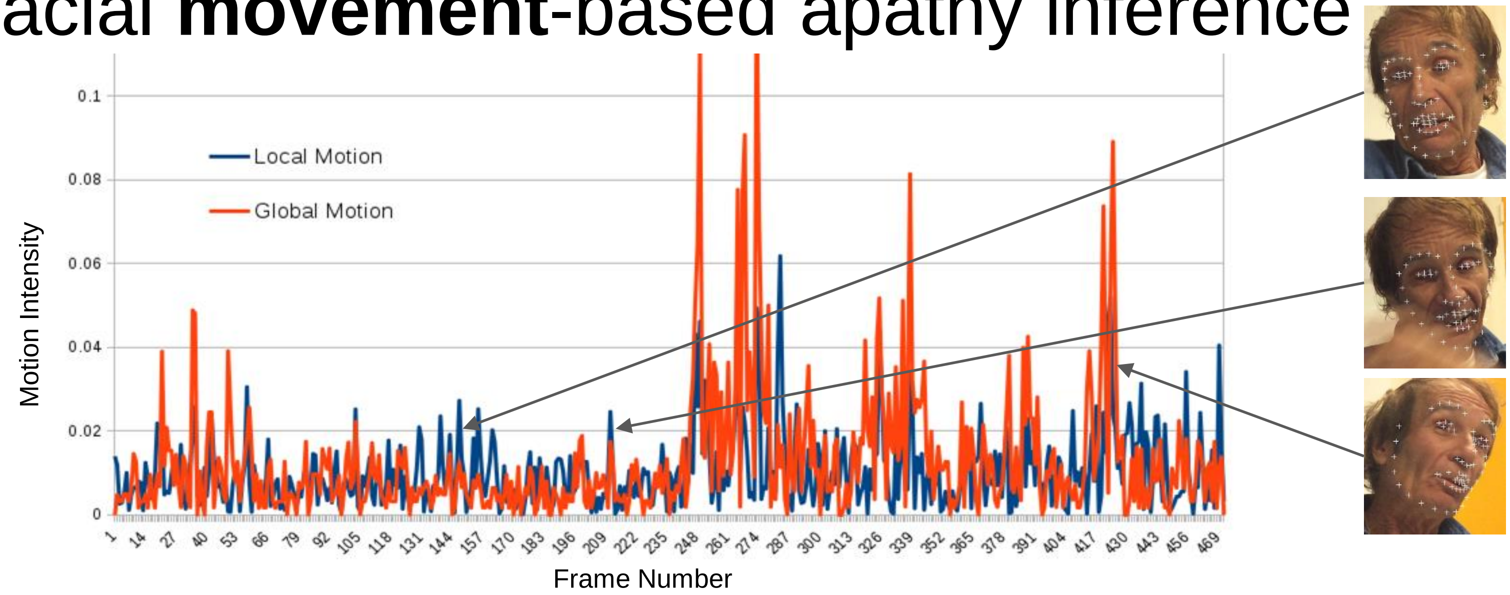


Negative narration



[Histograms of each emotion , Duration of each emotion , Histograms of each emotion , Duration of each emotion]

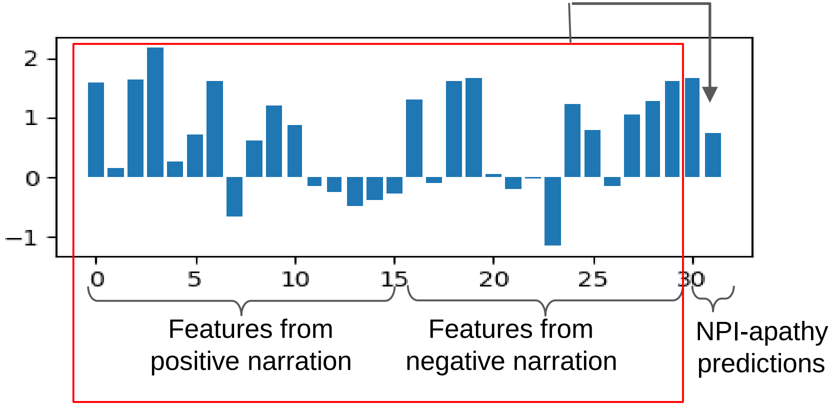
Facial **movement**-based apathy inference



- Global motion: head movement in successive frames
- Local motion: movement of facial parts, such as lips, eyebrows, cheeks, etc.
  - Computed by removing the global motion information from the interior facial landmarks

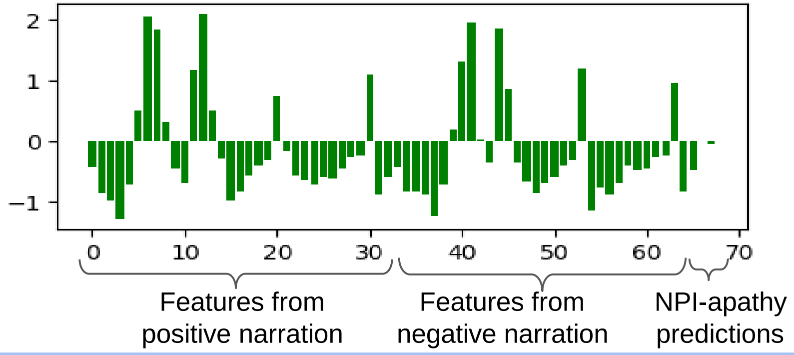
Emotion and Motion features

Motion features



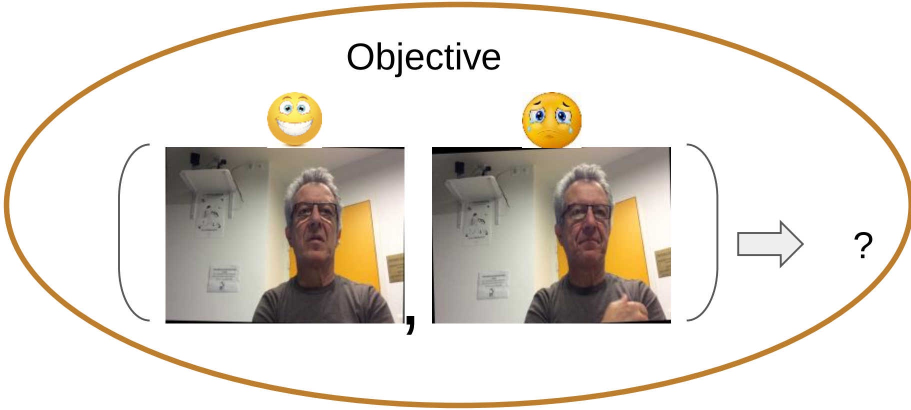
- *Movement of facial parts* in successive frames are represented in a histogram.

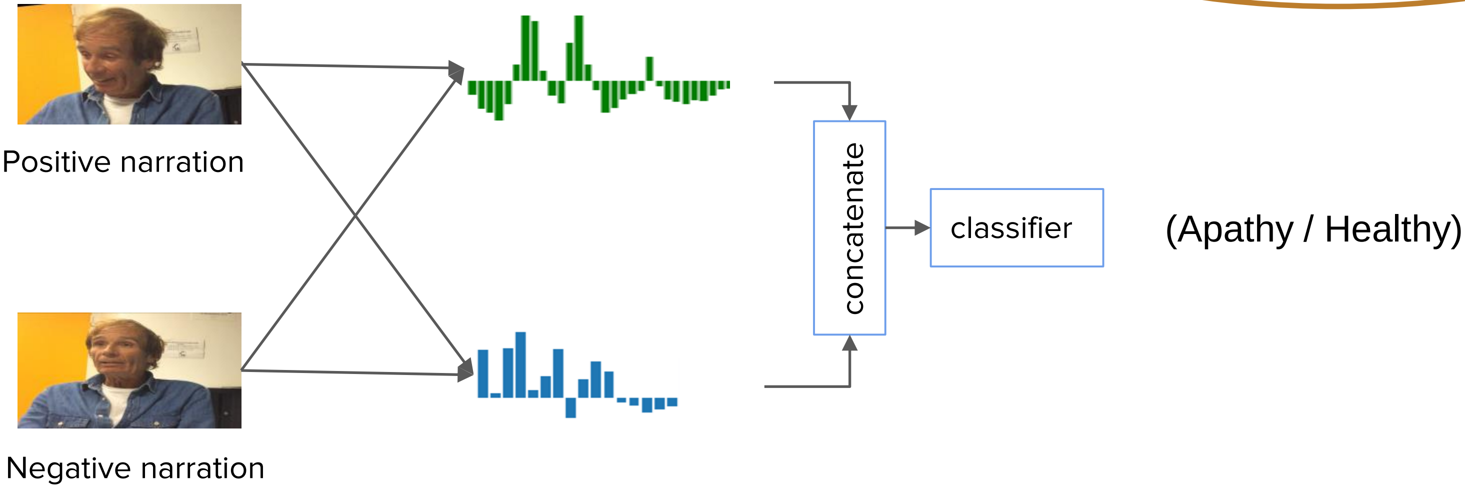
Emotion features



- The *expressions* in all frames are represented in a histogram.

Simplified Test Framework





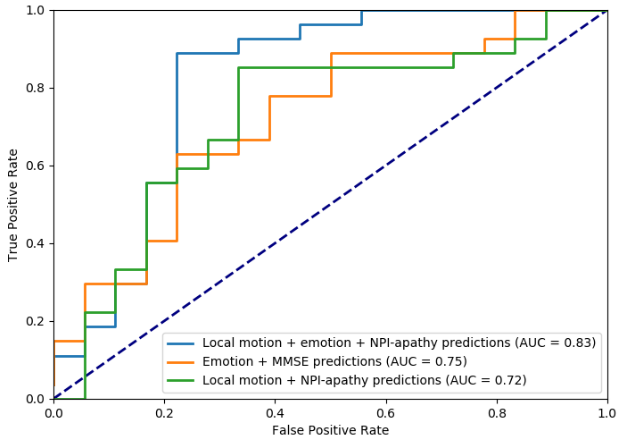
LOSO performances

Features used	*without* pos-neg concatenation			*with* pos-neg concatenation		
	Accuracy	F1-score	AUC	Accuracy	F1-score	AUC
Local Motion Features	58.88	0.505	0.527	57.77	0.555	0.558
Global Motion Features	56.66	0.516	0.523	55.55	0.537	0.537
Local + Global Motion Features	51.11	0.458	0.467	62.22	0.602	0.601
Emotion features	52.68	0.532	0.518	64.44	0.622	0.620
Emotion features + Local Motion Features	53.86	0.526	0.552	**77.77**	**0.757**	**0.75**
Emotion features + Local + Global Motion Features	54.57	0.532	0.537	71.11	0.68	0.676

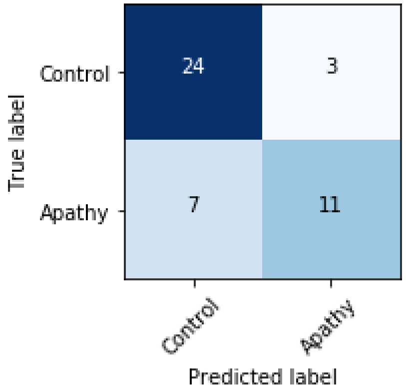
LOSO performances with pos-neg concatenation

Features used	Accuracy	F1-score	AUC
Local Motion Features + NPI-apathy prediction	73.33	0.722	0.722
Global Motion Features + NPI-apathy prediction	77.77	0.765	0.768
Local + Global Motion Features + NPI-apathy prediction	68.88	0.676	0.678
Emotion features + MMSE prediction	77.77	0.757	0.75
Emotion features + NPI-apathy prediction	66.66	0.649	0.648
Emotion features + Local Motion Features + MMSE prediction	68.88	0.675	0.675
Emotion features + Local Motion Features + NPI-apathy prediction	**84.44**	**0.836**	**0.833**
Emotion features + Local and Global Motion Features + MMSE prediction	68.88	0.660	0.657
Emotion features + Local and Global Motion Features + NPI-apathy prediction	77.77	0.757	0.75

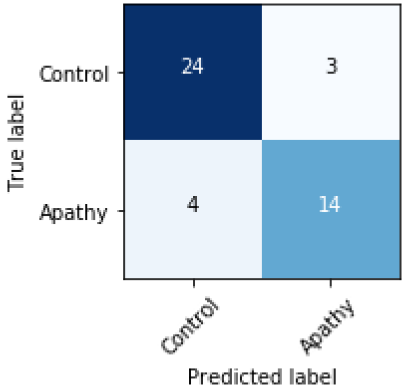
ROC and Confusion matrices



Motion and emotion features combined

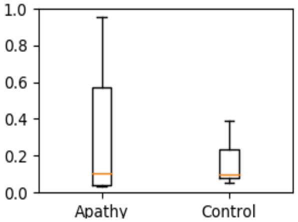


Motion and emotion features along with NPI-apathy prediction

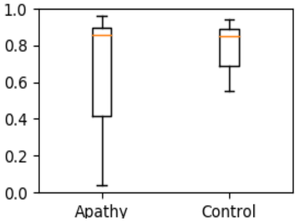


An observation

Male positive narration

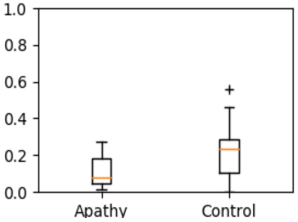


Negative emotion duration

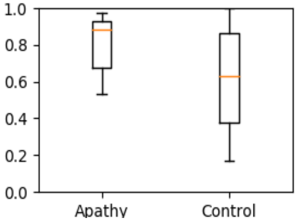


Neutral or positive emotion duration

Female negative narration



Negative emotion duration



Neutral or Positive emotion duration

- During positive narration, male subjects with Apathy showed both positive and negative emotions, whereas the control group showed more positive expressions compared to negative expressions.
- During negative narration, female subjects with Apathy showed less negative emotion compared to the control group.

Summary

- **First** to classify apathy based on **facial behavior** analysis
- Observation cues:
  - variation of facial **expressions**
  - facial **movements**
- **Regression** models to estimate the clinical scores
- Accuracy: 84%

To be presented at:

S L Happy, Antitza Dantcheva, Abhijit Das, Radia Zeghari, Philippe Robert, and Francois Bremond, "Characterizing the State of Apathy with Facial Expression and Motion Analysis," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2019 (FG-19).

Download here: https://goo.gl/zXFi38

Part 2: using Multi-task Learning

Multi-task learning (MTL)

[joint learning, learning to learn, and learning with auxiliary tasks]

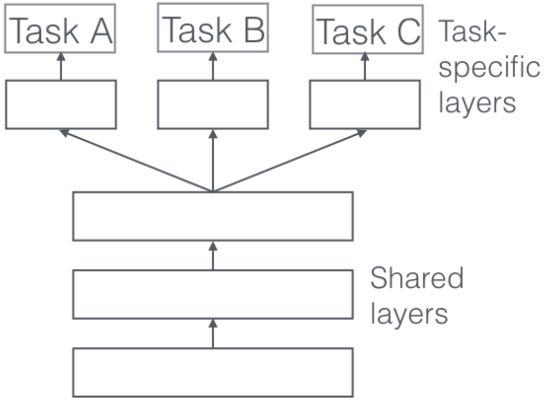
- MTL promotes sharing of model parameters to exploit the shared information across multiple tasks

Advantages:

- Knowledge transfer
- Learns general representation of all tasks
- Learns the relevant and irrelevant features for different tasks
- Joint learning results in a good regularization
- Ignores the data-dependent noise

Limitations:

- Negative transfer: sharing parameters with unrelated and dissimilar tasks



Experiment setting

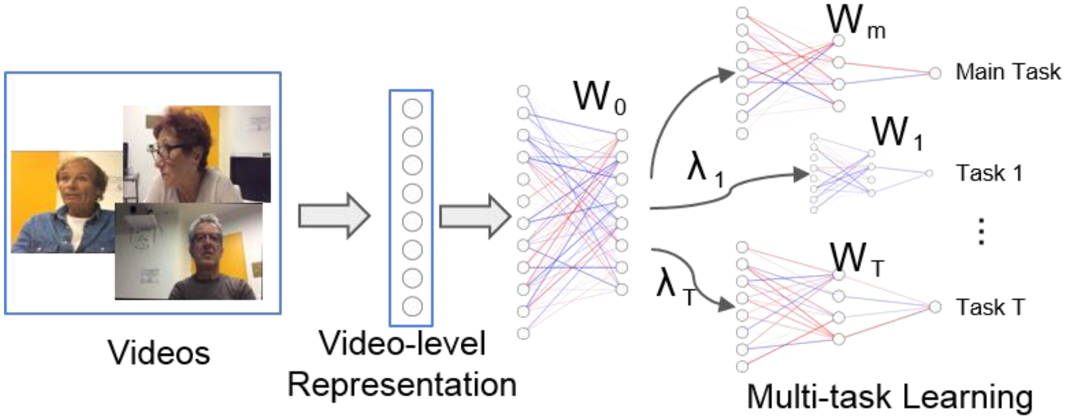
- Main task
  - Apathy classification

- Auxiliary regression tasks

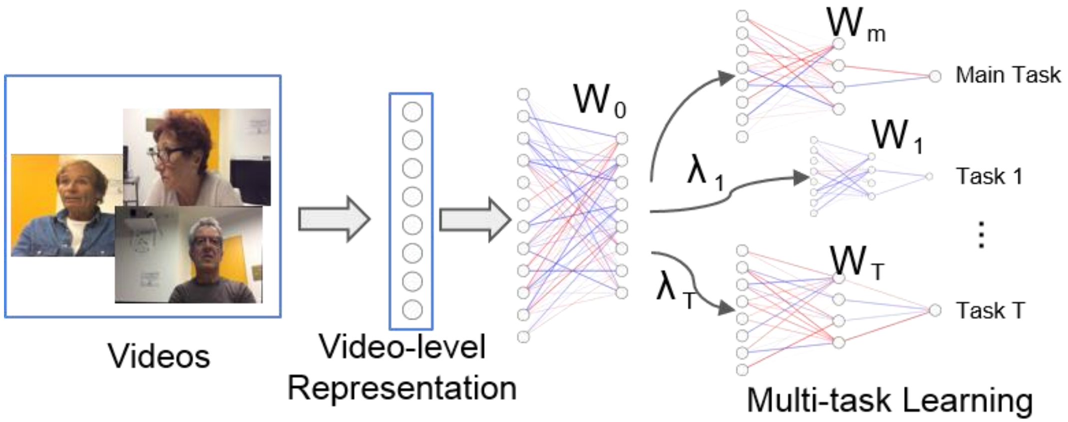
1. MMSE
2. NPI-apathy
3. NPI-anxiety
4. NPI-depression
5. NPI-total
6. clinical dementia score (CDR)
7. IA-affect
8. IA-initiative
9. IA-interest

NPI inventory

apathy inventory



Multi-task Learning



$$\arg\min_{\mathbf{W}_0,\mathbf{W}_m,\{\mathbf{W}_t\}_{a=1}^{T}} \sum_{i=1}^{N} \mathcal{L}(y_i^m, f(\mathbf{x}_i;\mathbf{W}_0,\mathbf{W}_m)) + \sum_{a=1}^{T}\sum_{i=1}^{N} \lambda^a \mathcal{L}(y_i^a, f(\mathbf{x}_i;\mathbf{W}_0,\mathbf{W}_a))$$

$\Rightarrow$

$$\arg\min_{\mathbf{W}} \quad \mathcal{L}^m + \sum_{a=1}^{T} \lambda^a \mathcal{L}^a$$

(Choose λ based on prior knowledge)

Objective

- MTL: $$\arg\min_{\mathbf{W}} \quad \mathcal{L}^m + \sum_{a=1}^{T} \lambda^a \mathcal{L}^a$$

- Our Objective (MTL+): $$\arg\min_{\mathbf{W}, \{\lambda^a\}_{a=1}^{T}} \quad \mathcal{L}^m + \sum_{a=1}^{T} \lambda^a \mathcal{L}^a$$

- (avoid negative transfer) learn the relatedness of the auxiliary tasks to the main task
- avoid trivial solution $\lambda^a = 0, \ \forall a$ (nullifying loss incurred by auxiliary tasks)

Proposed Method

- Initialize $W_{i,j} \sim U\left[-\sqrt{1/n}, \sqrt{1/n}\right]$ and $\lambda^a = 1, \forall a$
- Weight update by back propagation (several epochs)

$$\mathbf{W} \longleftarrow \mathbf{W} - \eta_1 \frac{\partial \mathcal{L}}{\partial \mathbf{W}}$$

- Penalize λ^a intermittently in the same manner

$$\lambda^a \longleftarrow \lambda^a - \eta_2 \mathcal{L}^a$$ (because $\frac{\partial \mathcal{L}}{\partial \lambda^a} = \mathcal{L}^a$)

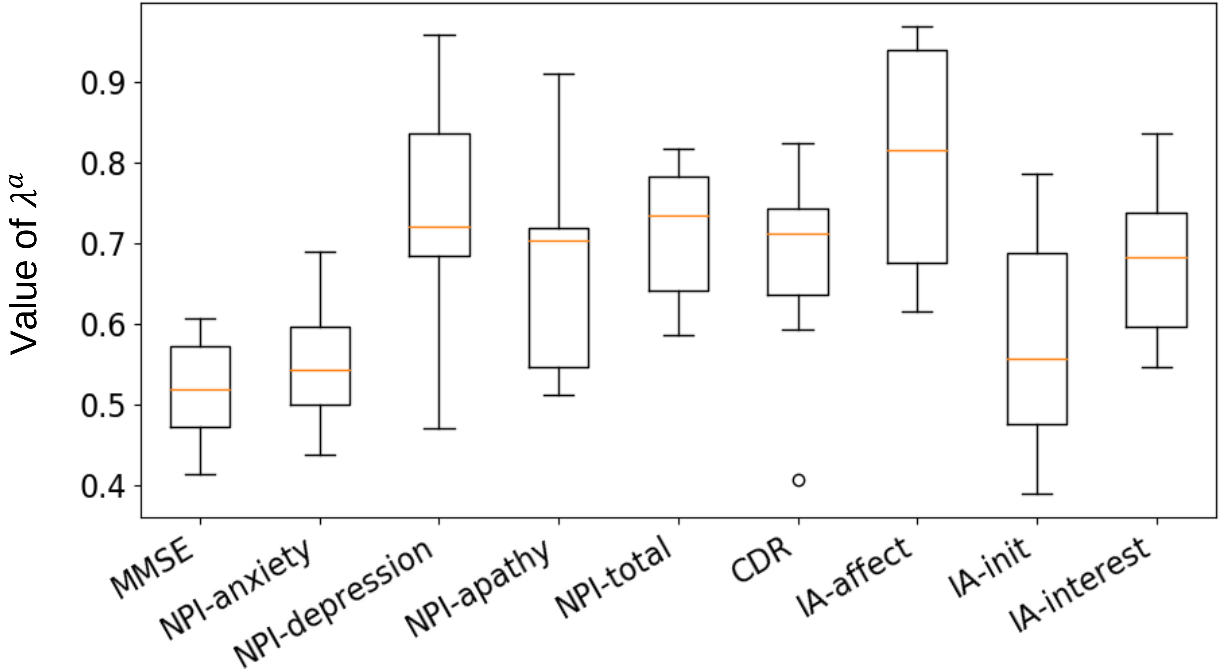
- Stopping criteria (to avoid over-fit the main task)

$$\frac{k.\text{med}_{j=t-k}^{t} E_{val}^{a}(j)}{\sum_{j=t-k}^{t} E_{val}^{a}(j) - k.\text{med}_{j=t-k}^{t} E_{val}^{a}(j)} > \epsilon$$

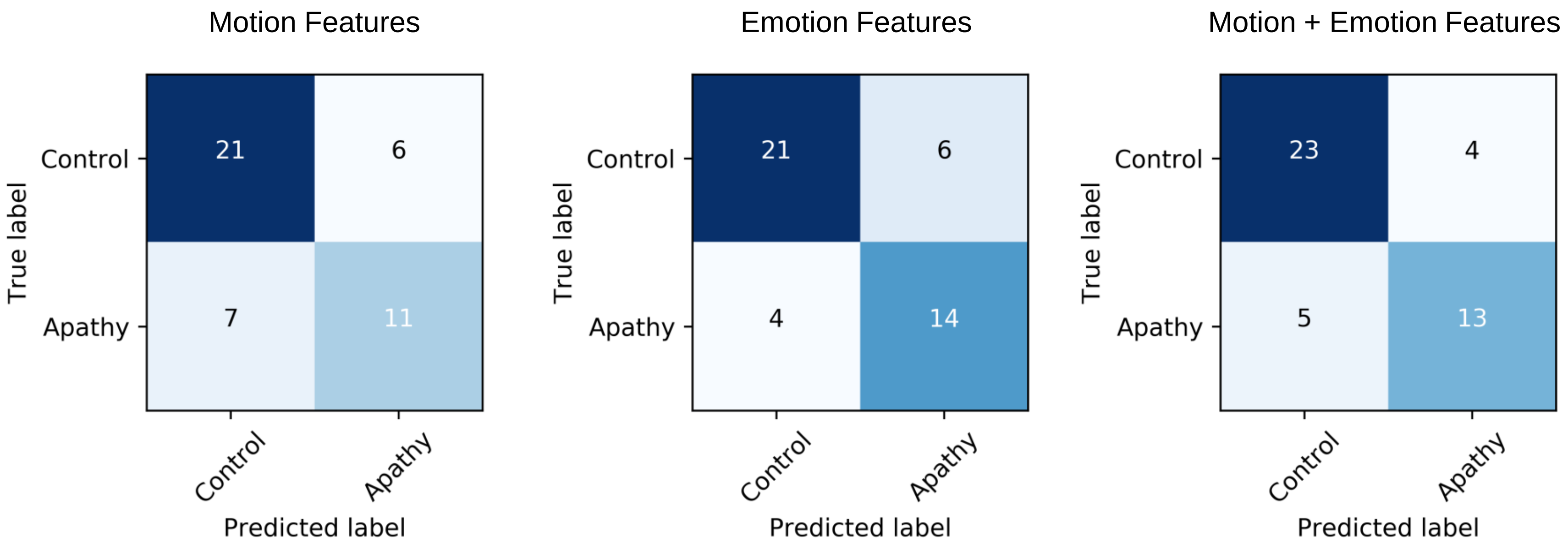
Performance

Features used	Accuracy	F1-score
MTL with Motion Features	62.22	0.582
MTL with Emotion features	66.66	0.638
MTL with Emotion + Motion Features	71.11	0.716
MTL+ with Motion Features	71.11	0.679
MTL+ with Emotion features	77.77	0.776
MTL+ with Emotion + Motion Features	**80.00**	**0.786**

How much are the tasks related?



Confusion Matrices



Summary

- Less accurate than the previous model
    - Probably due to less sample size, deep learning models could not learn well
    - Could improve the model performance with more data
- Learns the relatedness of different tasks/channels

Accuracy	80%
F1-score	0.786
Precision	0.825
Recall	0.816

Thank you