Multimodal Behaviour Modelling in Affective Computing Applications

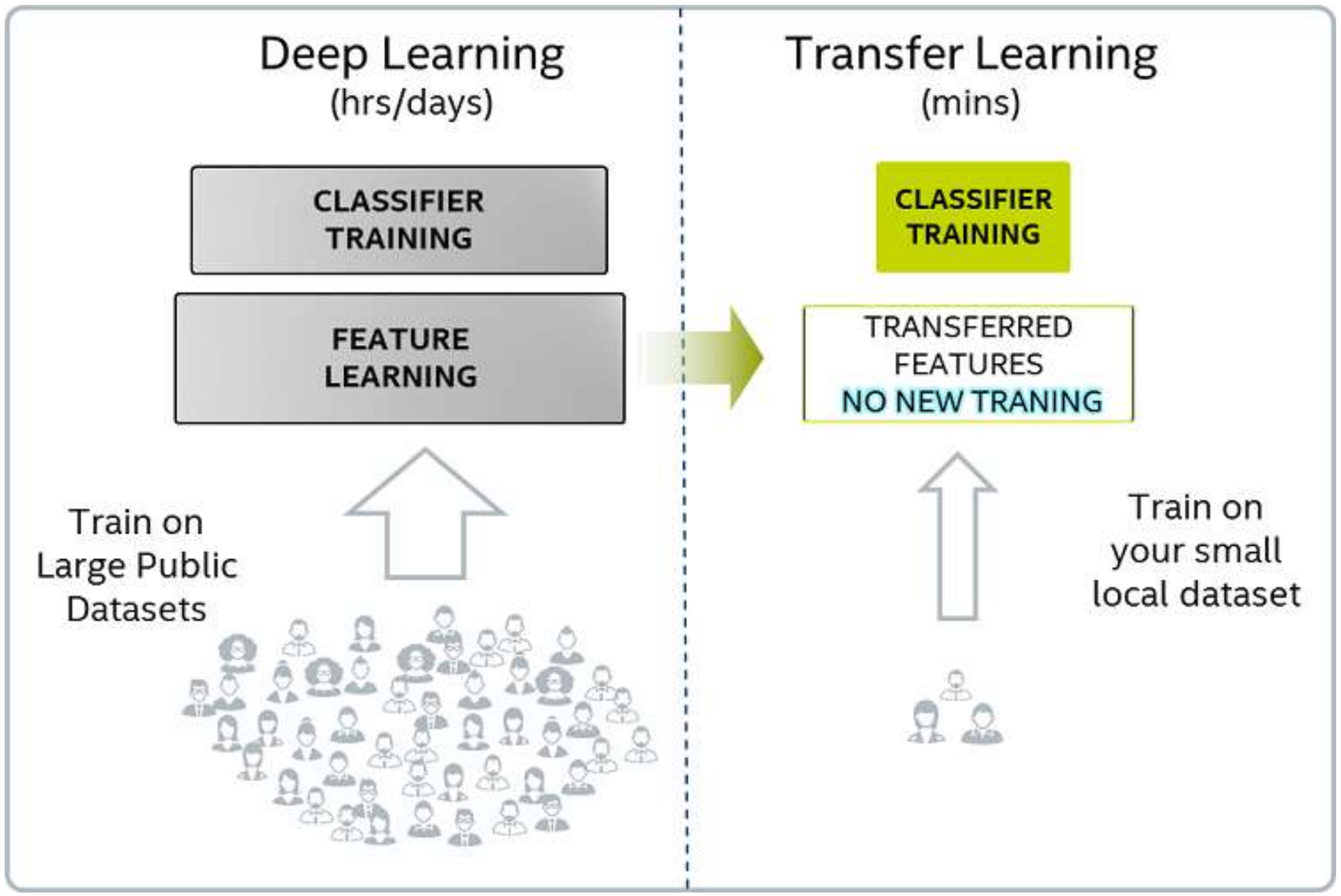
Marwa Mahmoud

Lecturer in Socially Intelligent Technologies,
University of Glasgow

Visiting Fellow, University of Cambridge

Transfer learning

-Transfer *learnt parameters from a related field* to a new field

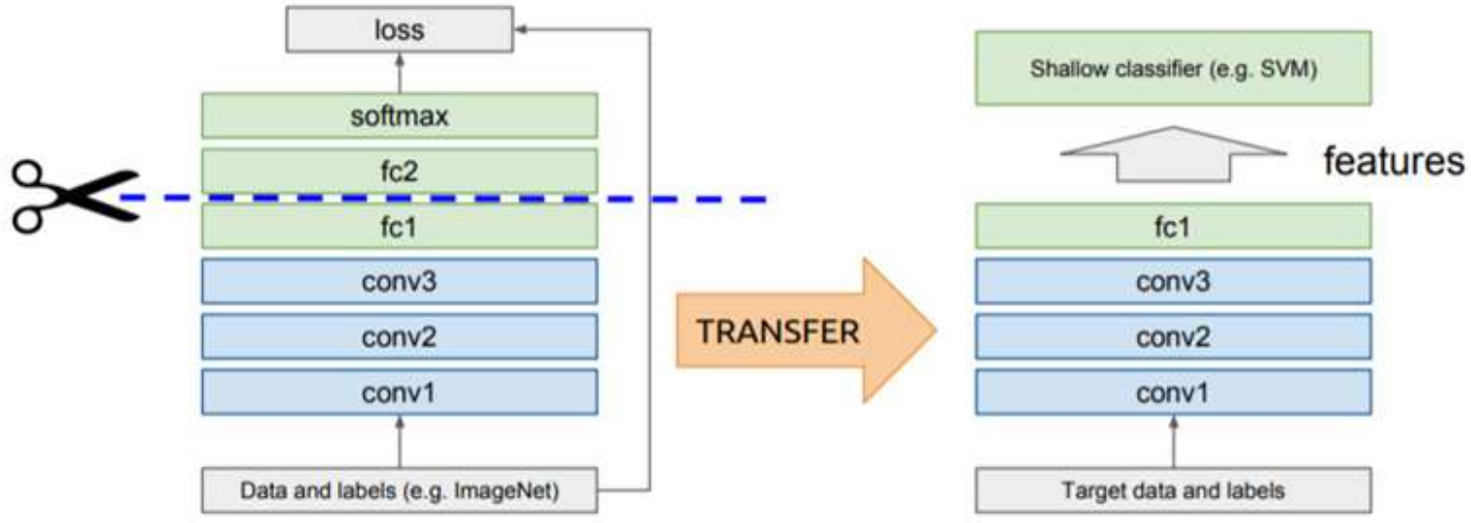


Transfer learning

Deep learning for feature extraction

Idea: use outputs of one or more layers of a network trained on a different task as generic feature detectors. Train a new shallow model on these features.

Assumes that $D_S = D_T$



Transfer learning – Fine-tuning

Freeze or fine-tune?

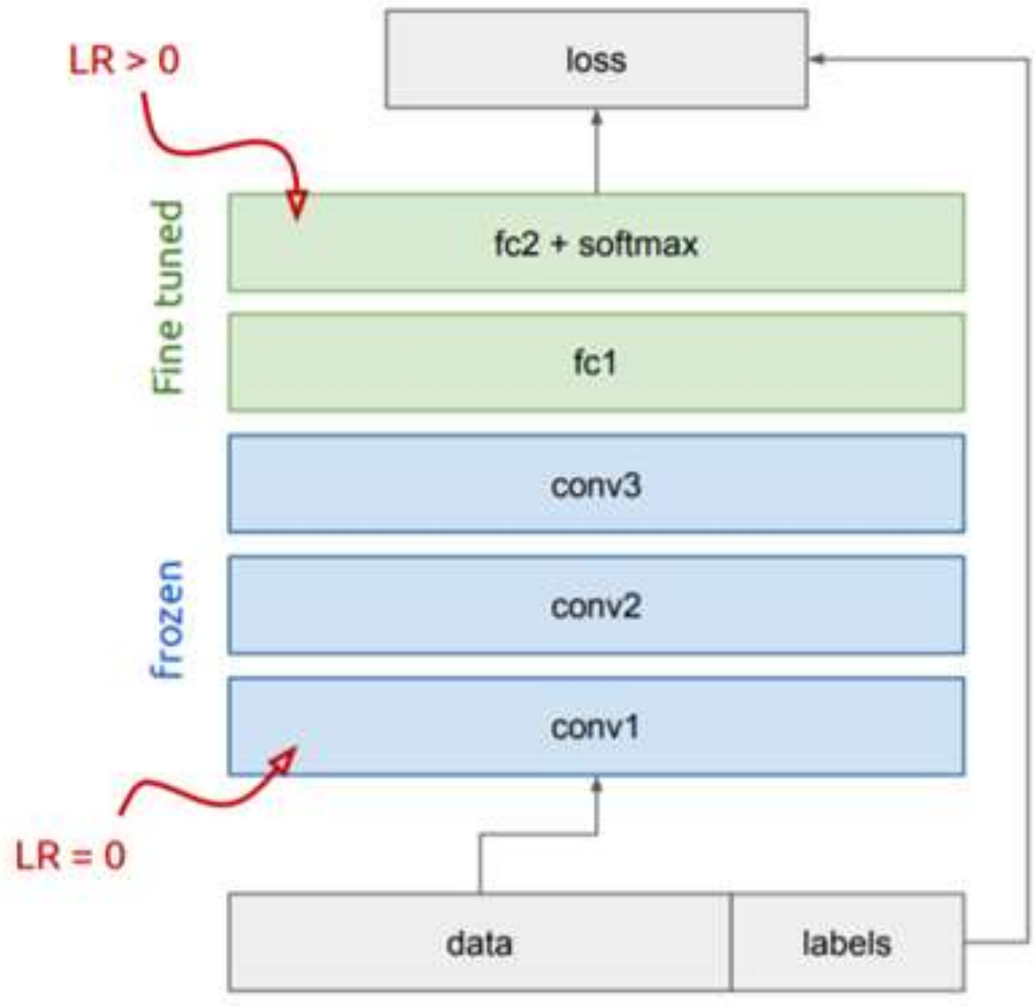
Bottom *n* layers can be frozen or fine tuned.

- **Frozen**: not updated during backprop
- **Fine-tuned**: updated during backprop

Which to do depends on target task:

- **Freeze**: target task labels are scarce, and we want to avoid overfitting
- **Fine-tune**: target task labels are more plentiful

In general, we can set learning rates to be different for each layer to find a tradeoff between freezing and fine tuning



Confusion matrix and evaluation

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{(\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative})}$$

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})}$$

Useful when the cost of false positives is high

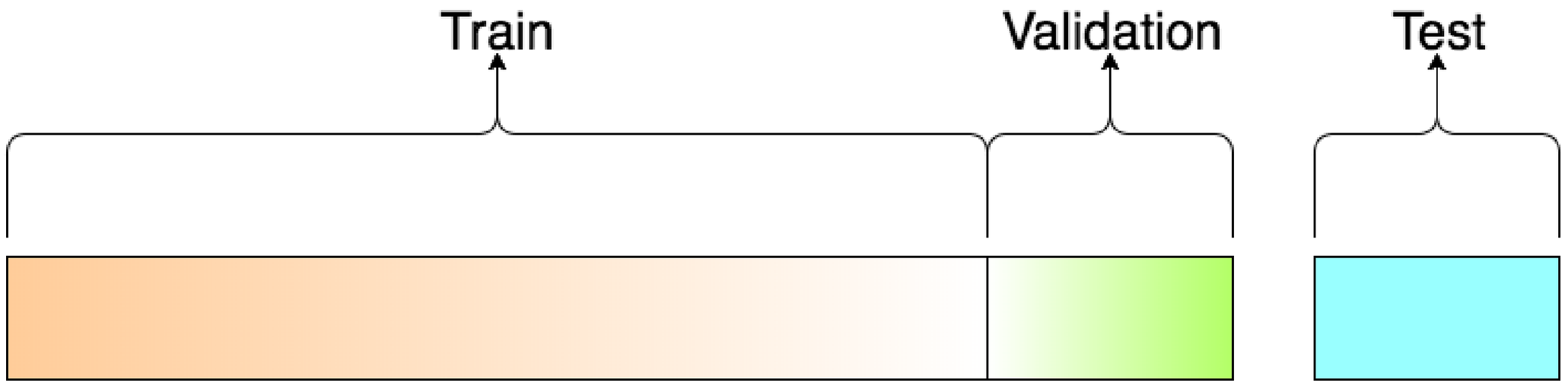
		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})}$$

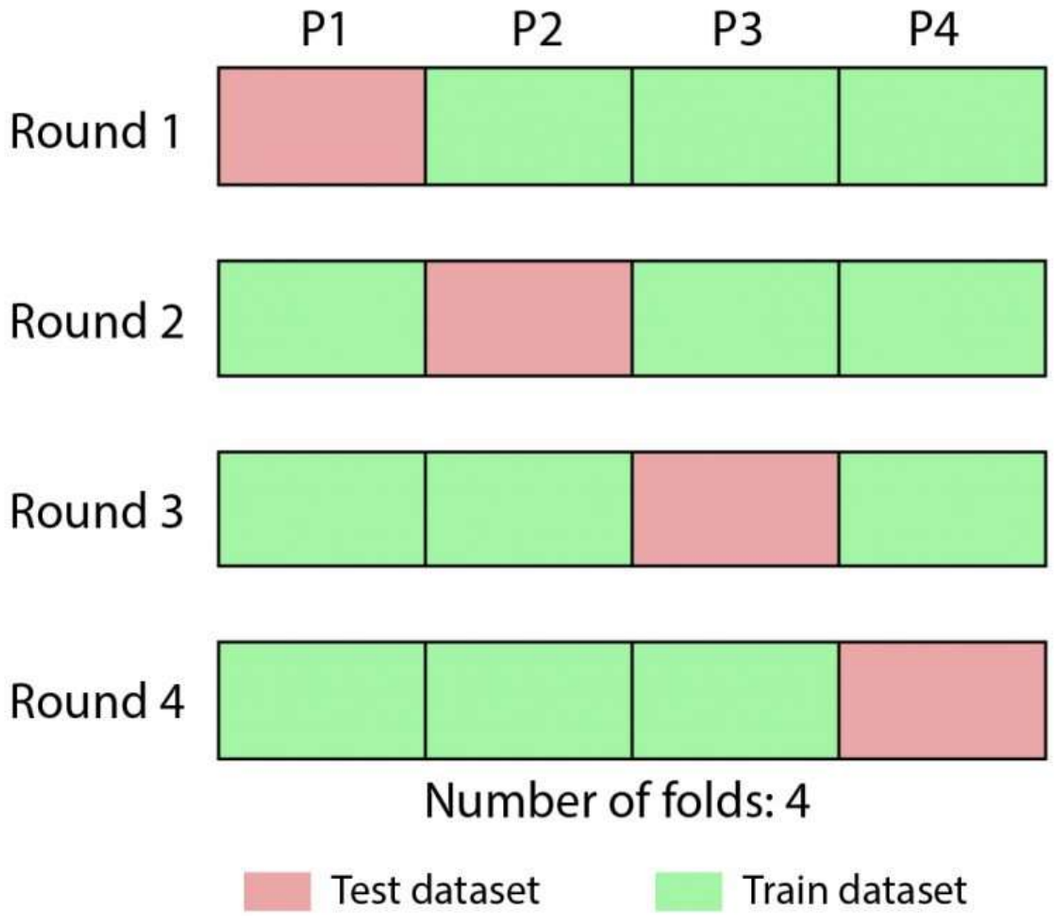
Useful when the cost of false negatives is high

$$\text{F1-score} = \left(\frac{\text{Recall}^{-1} + \text{Precision}^{-1}}{2}\right)^{-1} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

Datasets: training, validation, testing

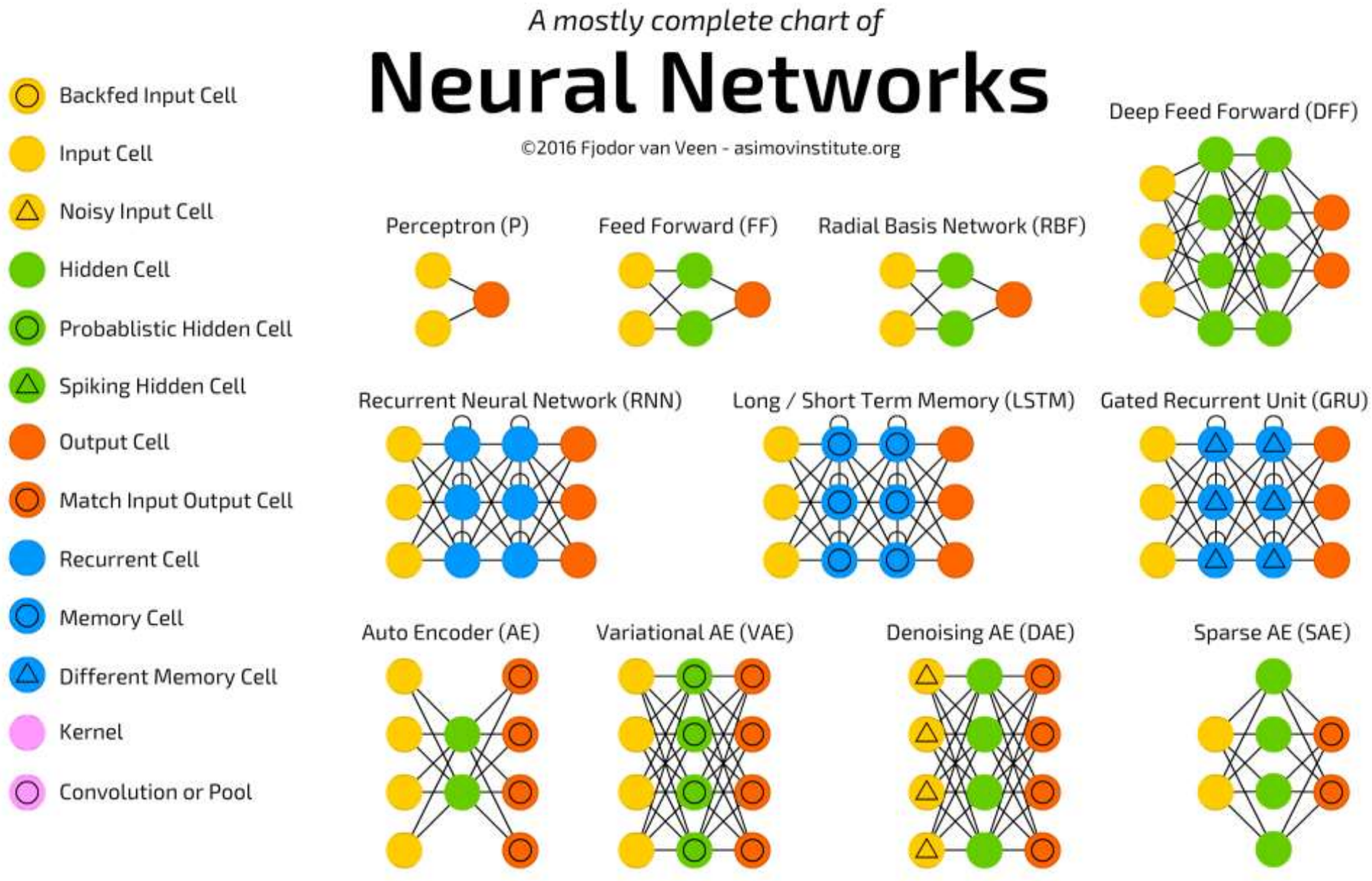


Cross validation



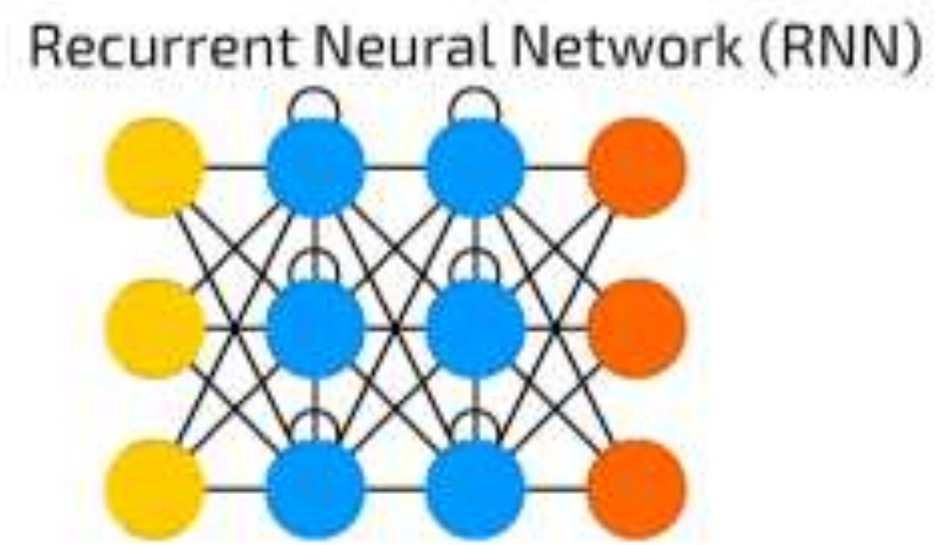
- **Participant- independent**
- **Optimise hyper-parameters using validation set (not training or testing)**
- **Data partitioning**

Deep Neural Networks



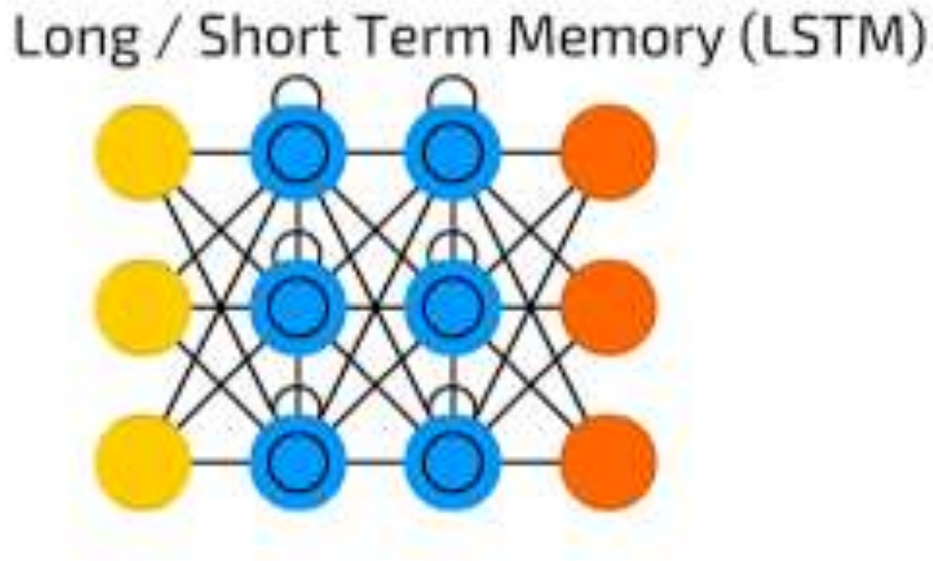
https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/neural_networks.html

Recurrent Neural Network



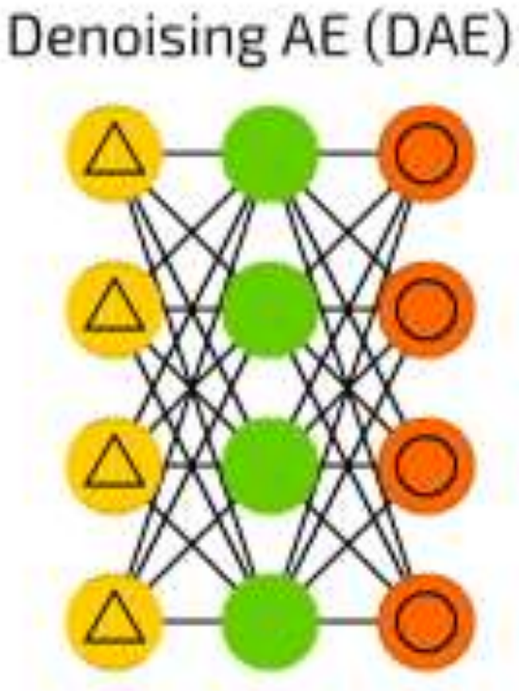
A **recurrent neural network** (RNN) is a class of artificial **neural networks** where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behavior. Unlike feedforward **neural networks**, RNNs can use their internal state (memory) to process sequences of inputs.

LSTM



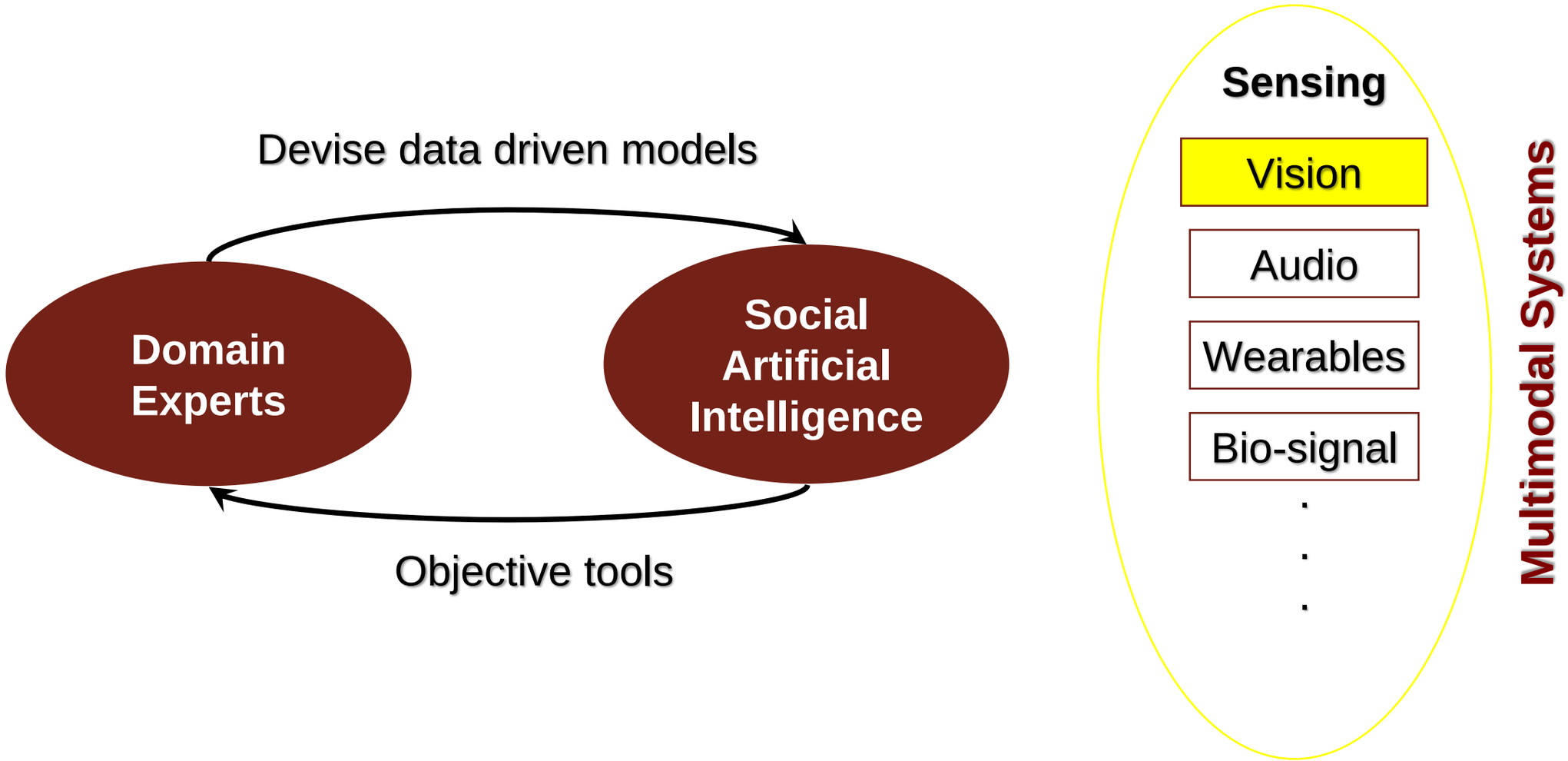
A **Long short-term memory (LSTM)** is an artificial Recurrent Neural Network (RNN) architecture that has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video).

Autoencoder



An **autoencoder** is a type of ANNs used to learn efficient data codings in an unsupervised manner. The aim of an autoencoder is to learn a representation (encoding) for a set of data, typically for dimensionality reduction, by training the network to ignore signal “noise”. Along with the reduction side, a reconstructing side is learnt, where the autoencoder tries to generate from the reduced encoding a representation as close as possible to its original input, hence its name.

Human behaviour understanding



Vision-based AI in Mental Health Support

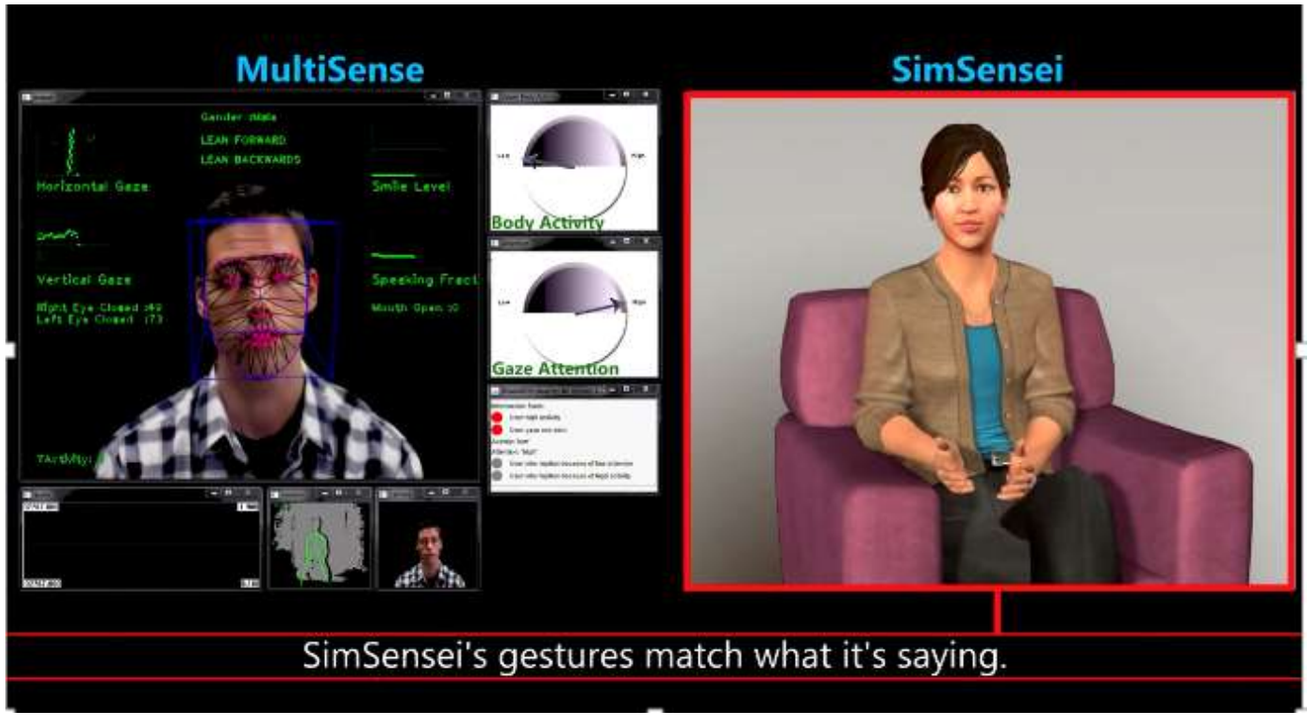
Psychological Distress

- Depression and anxiety are leading causes of disability that often go undetected or late-diagnosed

- Mental Health Foundation UK : *it is estimated that nearly two-thirds of people suffering distress have never received help from a health professional!*

- Automated detection of behavioural markers of distress: -
  - Objective tools
  - Complements self assessments
  - Supporting health professionals in decision-making (GP/early clinicians or lay health workers)

Audio-visual behaviour descriptors for psychological disorder analysis



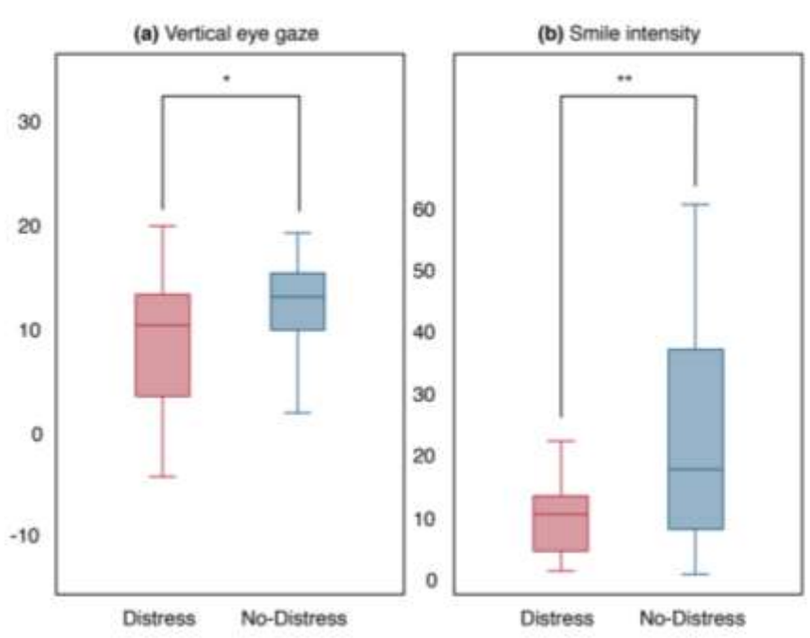
Authors	Nonverbal behavior	Disorder
Fairbanks, et al. 1982	↓ mouth movements	depression
	↓ *smiling*	
	↑ *self-grooming*	
	↑ *turning head away*	
	↑ *fidgeting*	anxiety
Hall, et al. 1995	↓ gestures	depression
	↓ speech	
	↑ long pauses	
Kirsch and Brunnhuber 2007	↑ anger	PTSD
	↓ *genuine joy*	
Perez and Riggio 2003	↑ *gaze down*	depression
	↑ *gaze aversion*	
	↓ emotional expressivity	
	↓ gestures	
	↑ frowns	
Schelde 1998	↑ *nonspecific gaze*	depression
	↓ mouth movements	
	↓ interaction	
Waxer 1974	↓ *mutual gaze*	depression

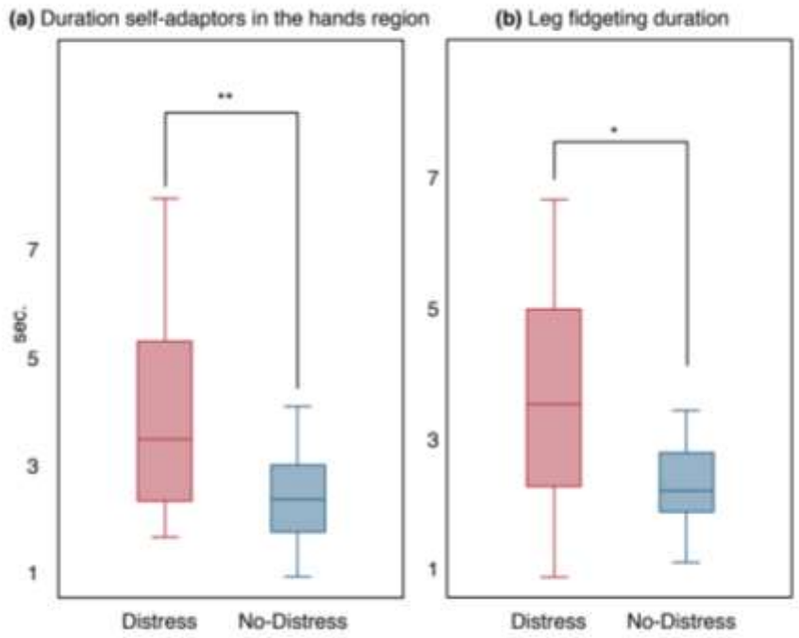
TABLE I

SUMMARY OF NONVERBAL BEHAVIORS FOUND IN THE LITERATURE. NONVERBAL BEHAVIORS WRITTEN IN ITALICS ARE PART OF THE ANALYSIS IN THE PRESENT WORK.

[Automatic audiovisual behavior descriptors for psychological disorder analysis. Stefan Scherer, Giota Stratou, Gale Lucas, Marwa Mahmoud, Jill Boberg, Jonathan Gratch, Albert (Skip) Rizzo and Louis-Philippe Morency. in Image and Vision Computing Journal, 2014]

Audio-visual behaviour descriptors for psychological disorder analysis





[Automatic audiovisual behavior descriptors for psychological disorder analysis. Stefan Scherer, Giota Stratou, Gale Lucas, Marwa Mahmoud, Jill Boberg, Jonathan Gratch, Albert (Skip) Rizzo and Louis-Philippe Morency. in Image and Vision Computing Journal, 2014]

Well-being dataset

- Available datasets:
  - Clinical (private)
  - Non-clinical (based on self-report)

(Only features are shared)

[Indigo Ortan & Gabriela Pavarini, Marwa Mahmoud. ‘Vision based body gesture meta features for Affective Computing’,2019]

Automatic detection of fidgeting

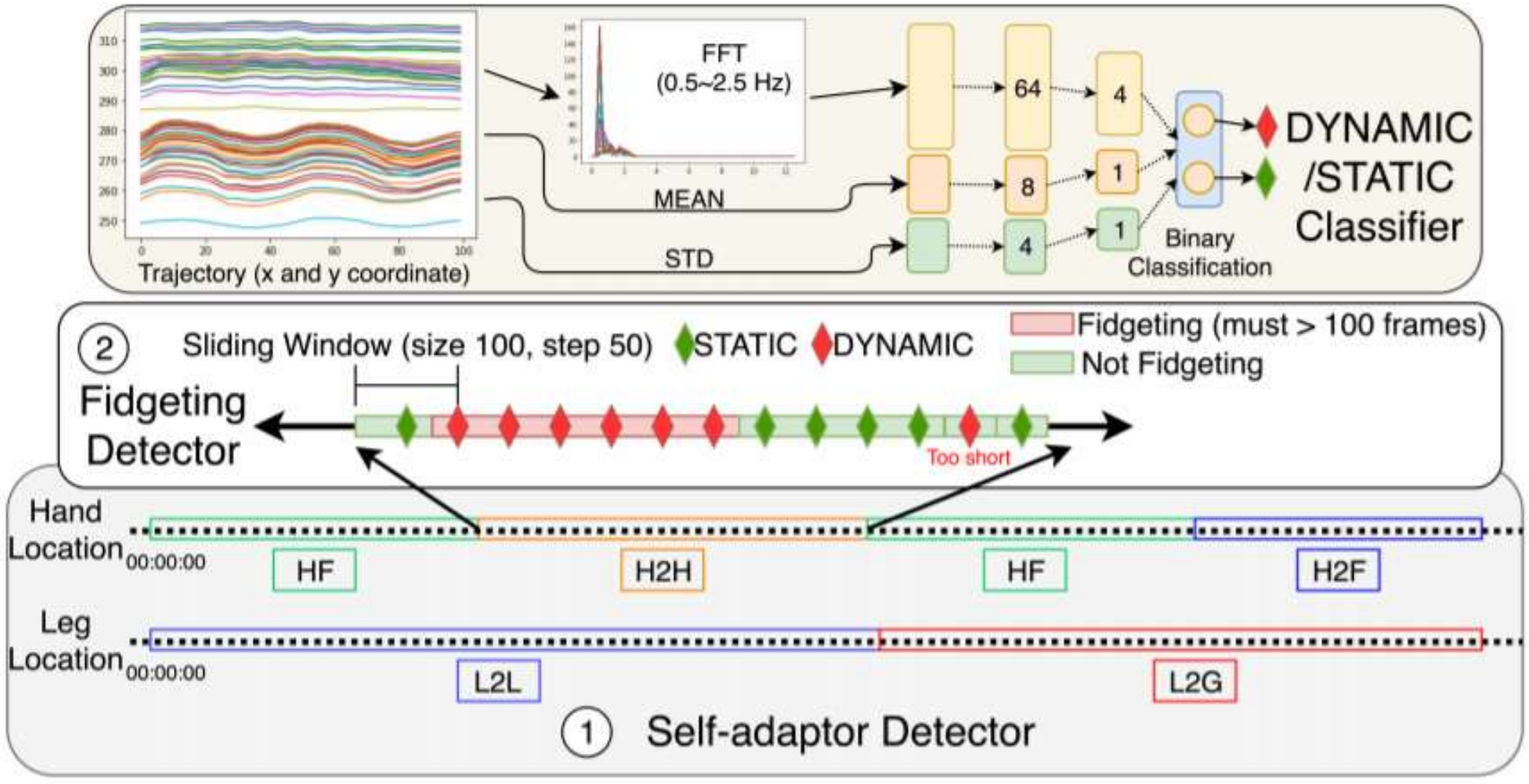
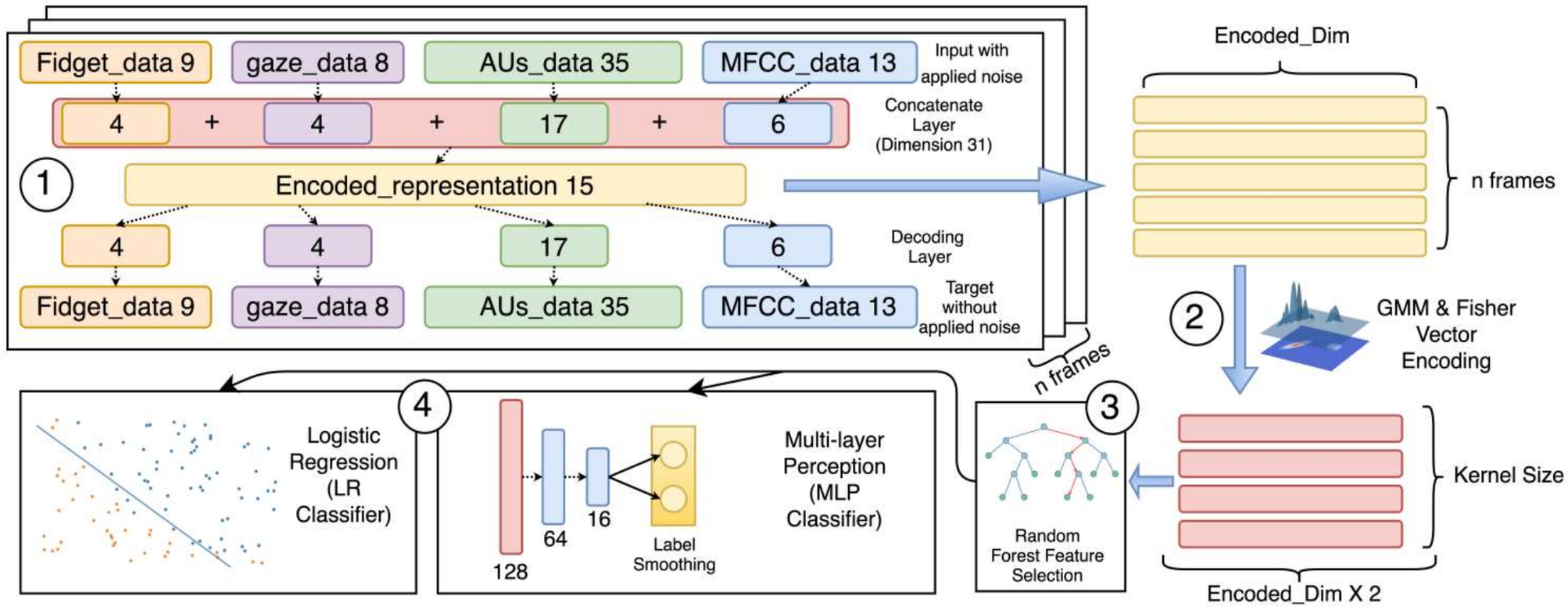


Fig. 1. Hierarchical self-adaptor detection workflow. (1) First detect hand/leg location (2) Classify motion using *DYNAMIC/STATIC Classifier* and then finally combine location and motion to give high-level fidgeting event. Figure shows the detection of H2H (Hand to hand) fidget. Same principle applies to other fidgets.

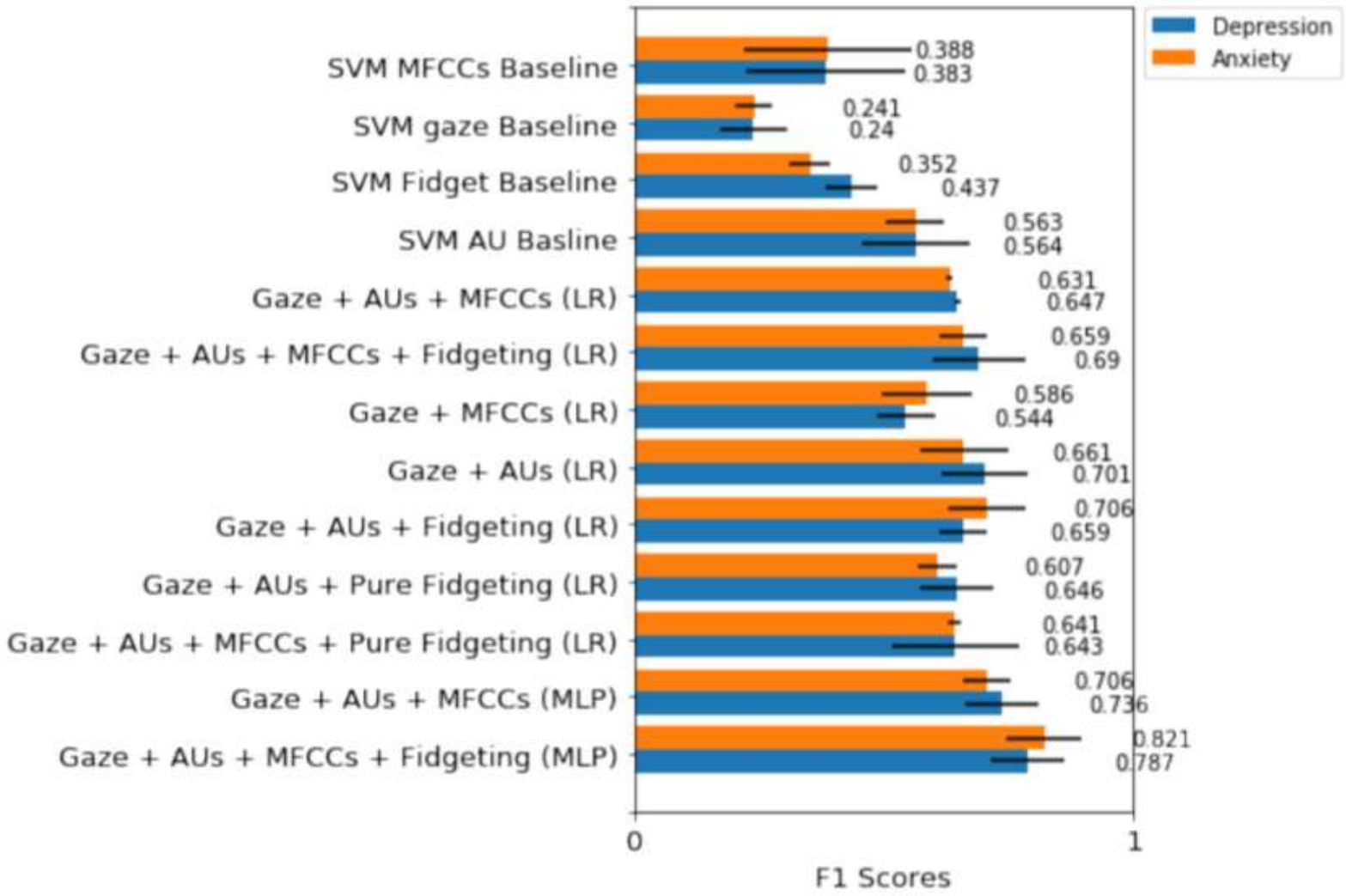
[Weizhe Lin, Indigo Orton, Mingyu Liu, and Marwa Mahmoud. 'Automatic Detection of Self-Adaptors for Psychological Distress', FG 2020]

Automatic detection of psychological distress: multimodal framework



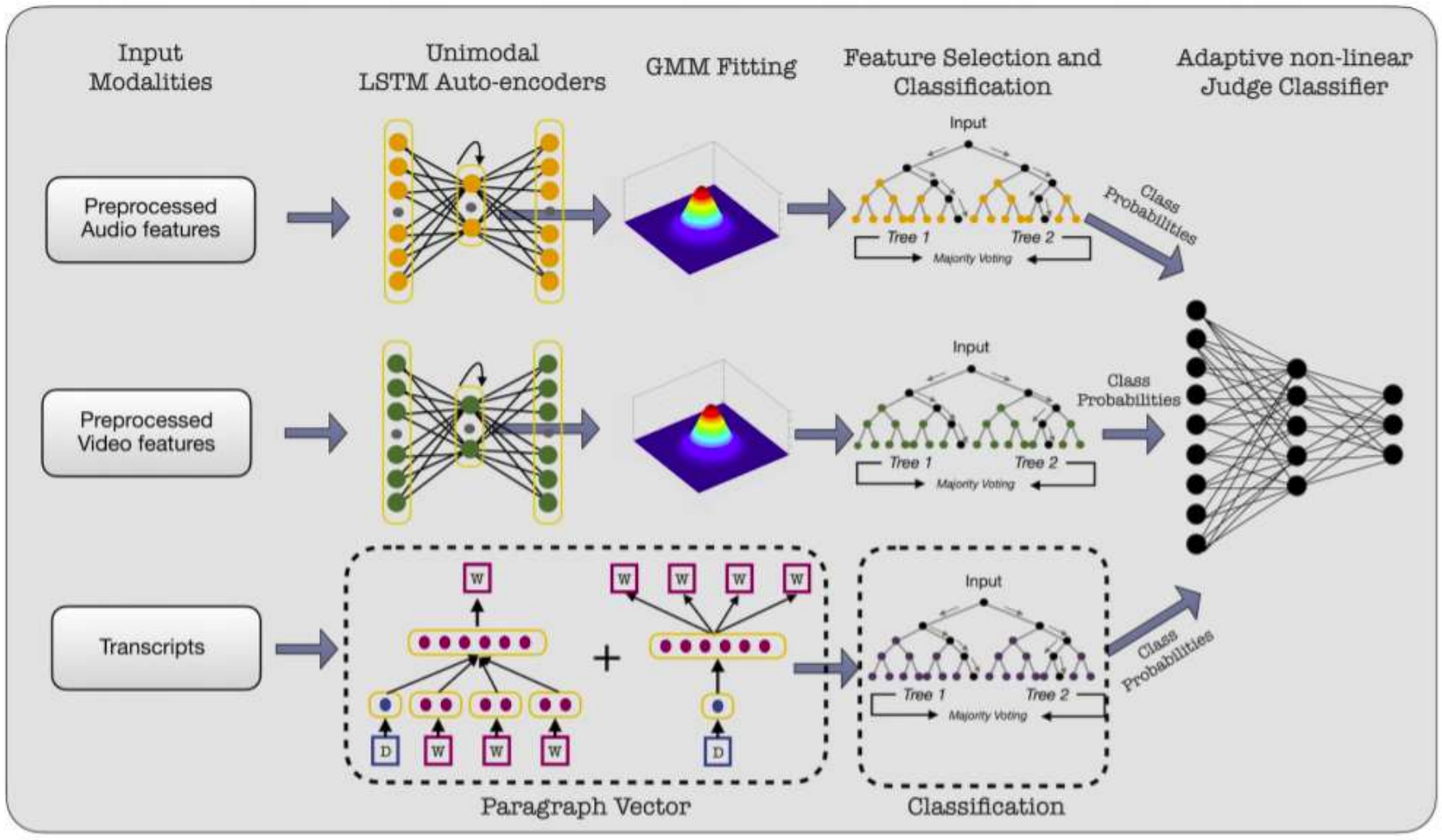
[W. Lin, I Orton, Q. Li, G. Pavarini and M. Mahmoud. “Looking At The Body: Automatic Analysis of Body Gestures and Self-Adaptors in Psychological Distress”. IEEE Transaction on Affective Computing, 2021

Automatic detection of psychological distress: multimodal framework



[W. Lin, I Orton, Q. Li, G. Pavarini and M. Mahmoud. “Looking At The Body: Automatic Analysis of Body Gestures and Self-Adaptors in Psychological Distress”. IEEE Transaction on Affective Computing, 2021

Multimodal temporal machine learning models



The audio and visual modalities are encoded using bidirectional LSTM models. The descriptors for the whole videos are generated using Fisher Vector. For the textual modality, Paragraph-Vector is proposed. A final deep classifier is used to combine the unimodal predictions.

[Francesco Ceccarelli and Marwa Mahmoud. 'Multimodal temporal machine learning for Bipolar Disorder and Depression Recognition', Pattern Analysis and Applications Special Issue on Computer Vision and Machine Learning for Healthcare Applications, 2021]

Multimodal temporal machine learning models

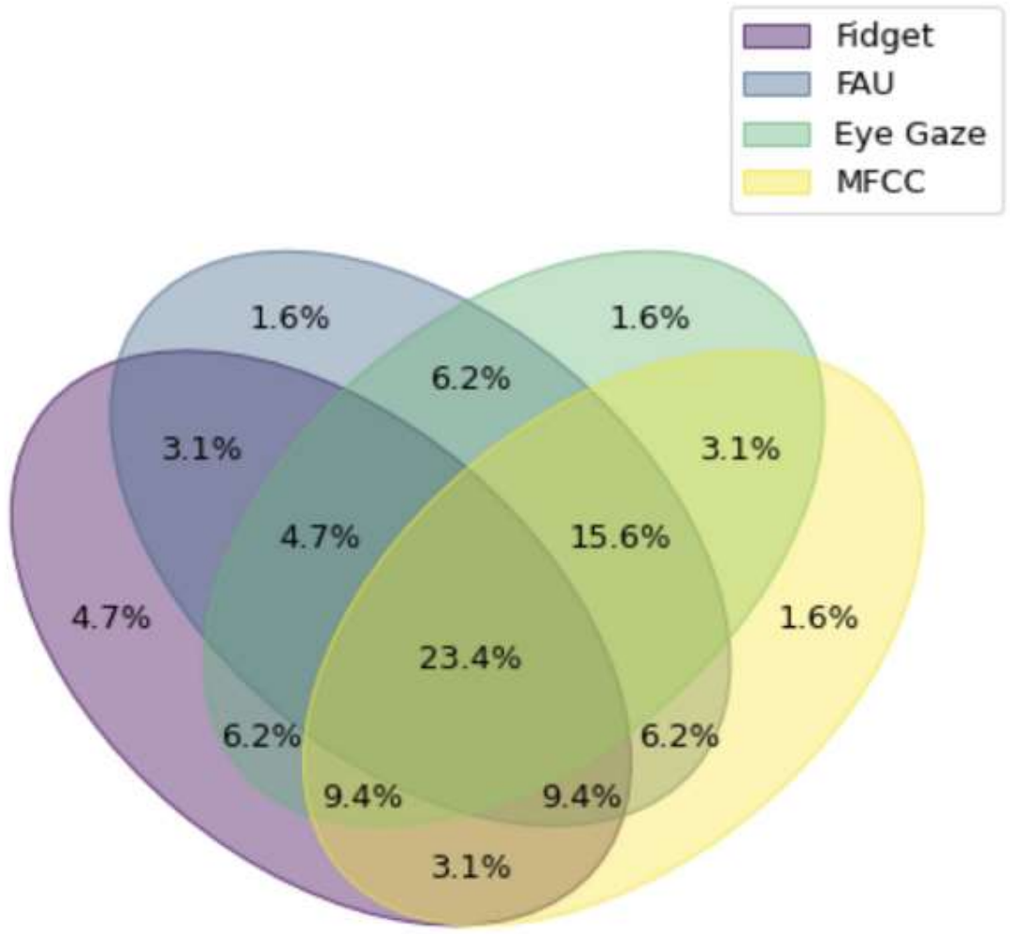
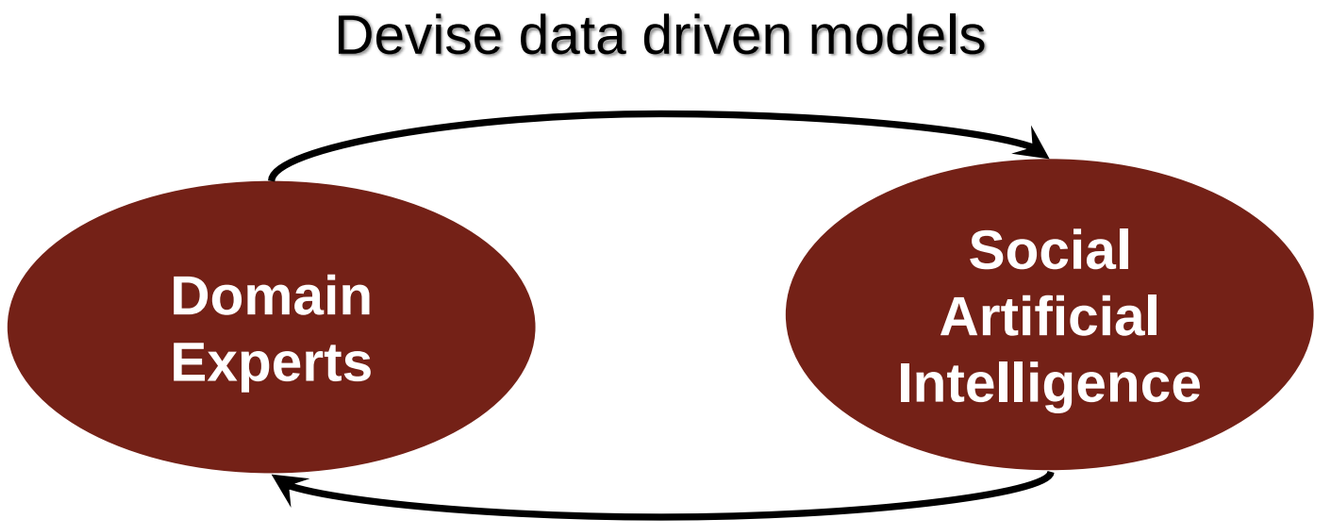


Fig. 3 Venn diagram for modality importance. Each modality captures useful information that other modalities fail to associate to mental disorder (i.e each modality has a non-overlapping percentage of samples which are correctly classified by that modality only).

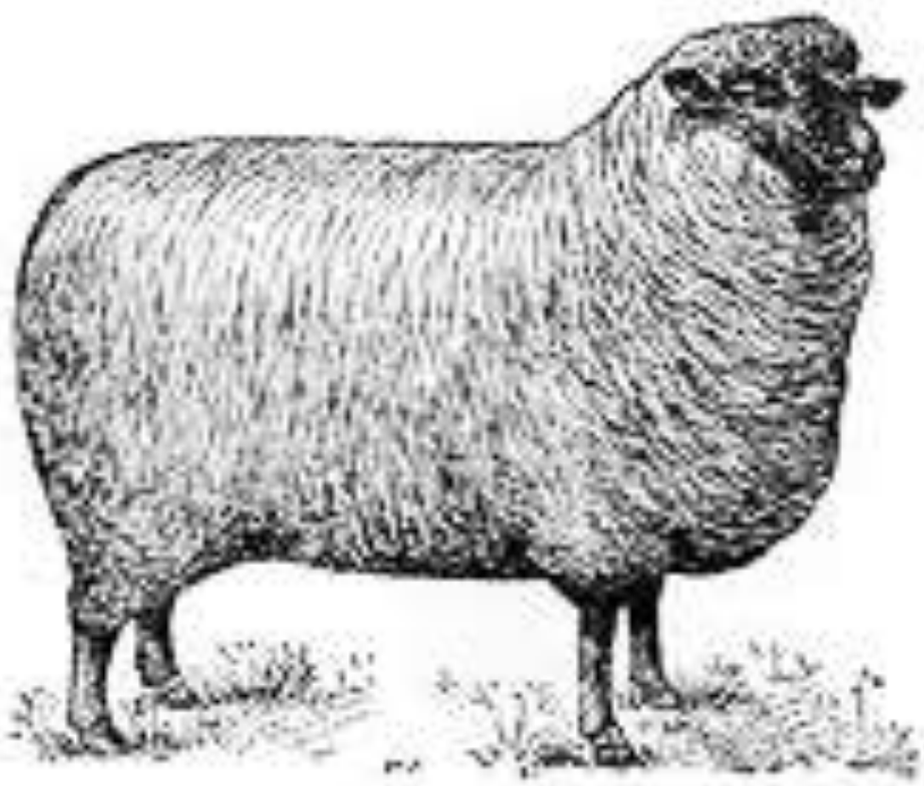
[Francesco Ceccarelli and Marwa Mahmoud. 'Multimodal temporal machine learning for Bipolar Disorder and Depression Recognition', Pattern Analysis and Applications Special Issue on Computer Vision and Machine Learning for Healthcare Applications, 2021]

Behaviour modelling for mental health

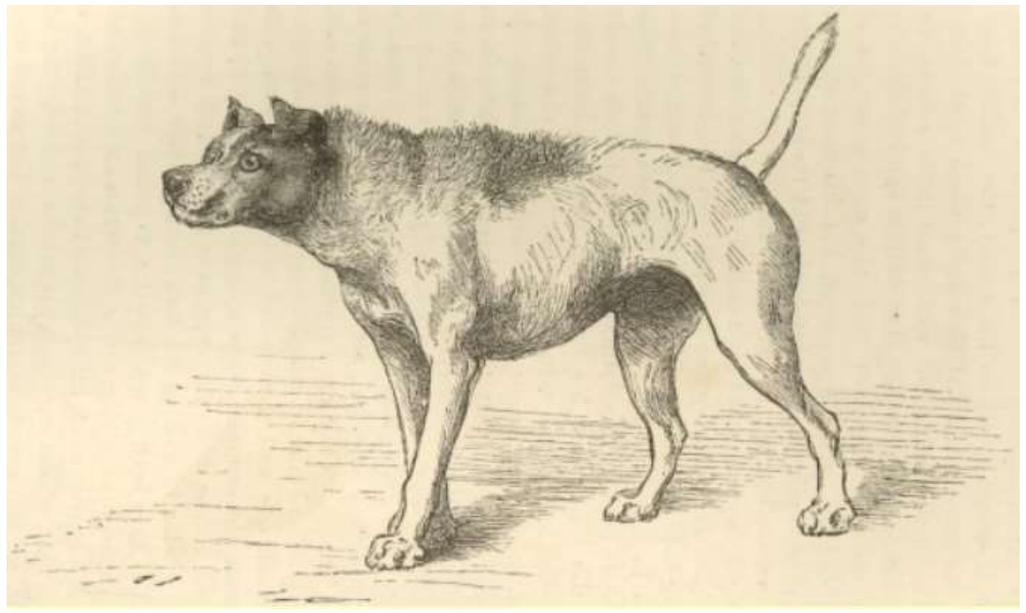


- Automatic detection of novel mid-level features
- ML models: Noisy, multidimensional , usually small, datasets
- Probabilistic models as first line of support

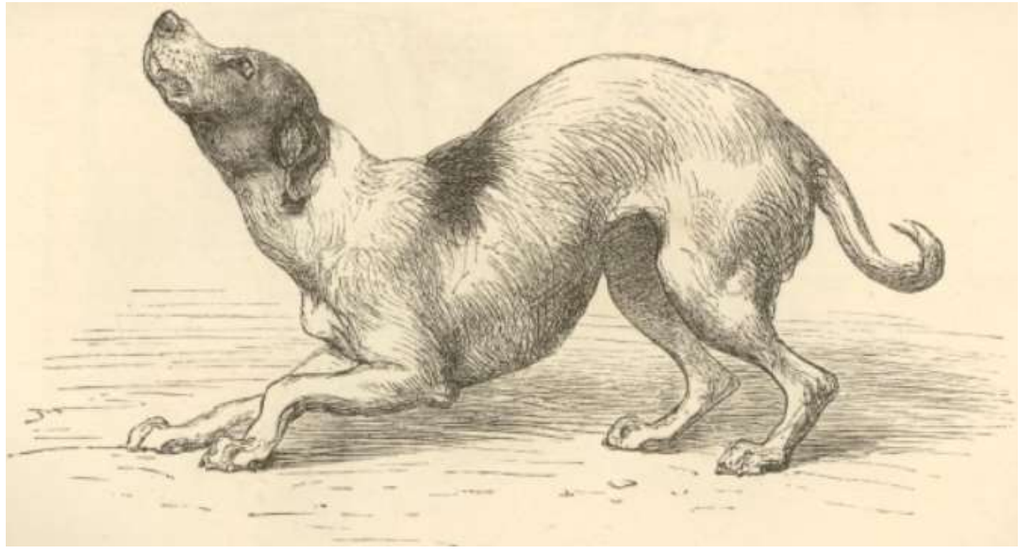
Vision-based AI in Animal Welfare



The expression of the emotions in man and animals, *Darwin 1872*



Dog approaching another dog with hostile intentions. By Mr. Riviere.



The Same in a humble and affectionate frame of mind. By Mr. Riviere.

The same, when pleased by being caressed.

The Sheep Pain Facial Expression Scale (SPFES)

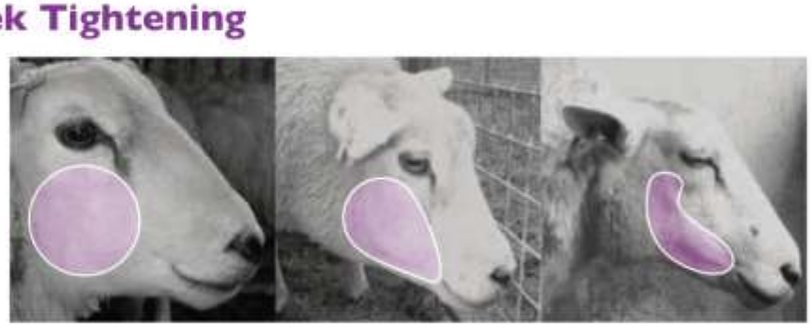
Tightening of the eye

Not present = 0 | Partially present = 1 | Present = 2

Abnormal ear position - front view

Not present = 0 | Partially present = 1 | Present = 2

Cheek Tightening



Not present = 0 | Partially present = 1 | Present = 2

Abnormal ear position - profile view

Not present = 0 | Partially present = 1 | Present = 2

Abnormal lip and jaw profile

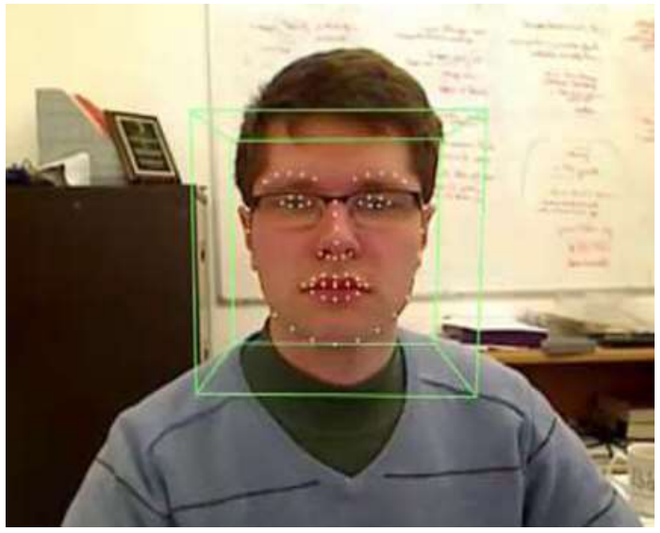
Not present = 0 | Partially present = 1 | Present = 2

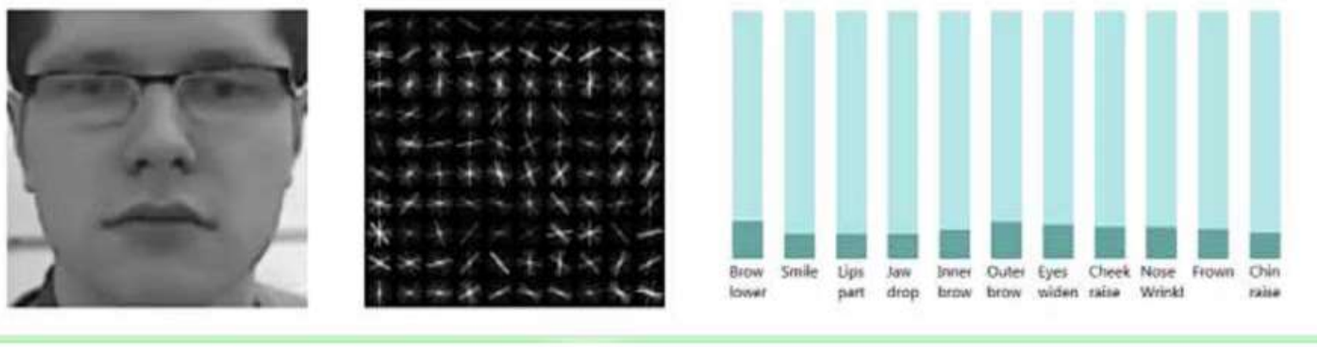
Abnormal nostril and philtrum shape

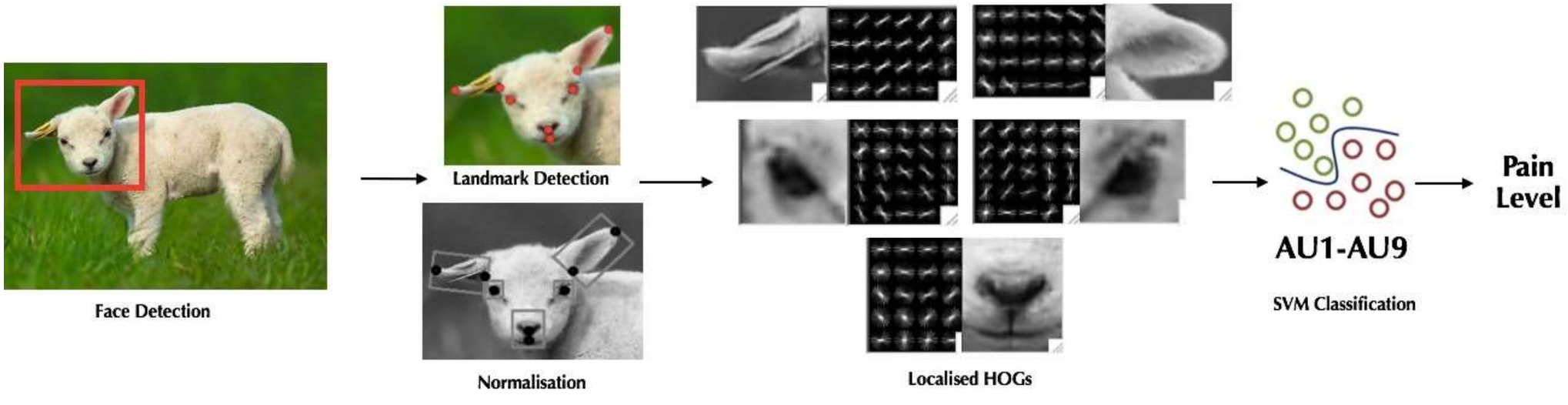
Not present = 0 | Partially present = 1 | Present = 2

McLennan, K.M., et al., Development of a facial expression scale using footrot and mastitis as models of pain in sheep. Applications on Animal Behaviour Science, 2016.

Automatic detection of facial expressions of animals pipeline



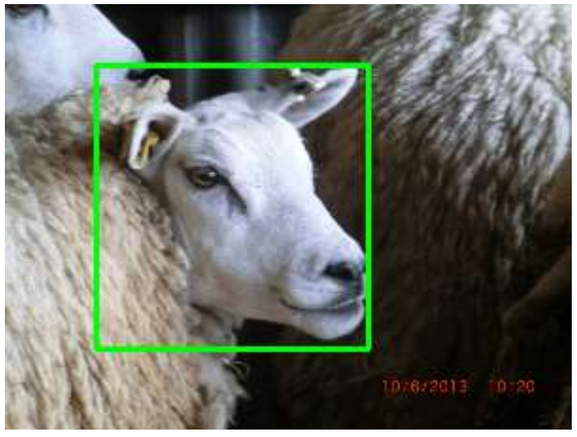
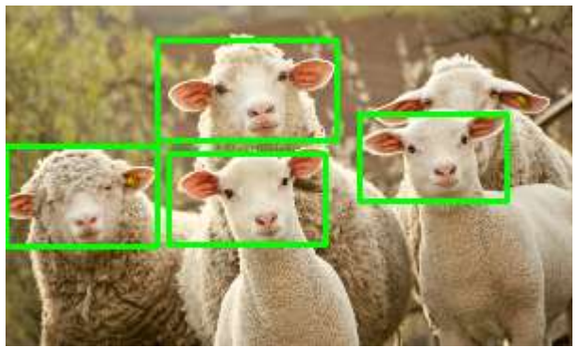
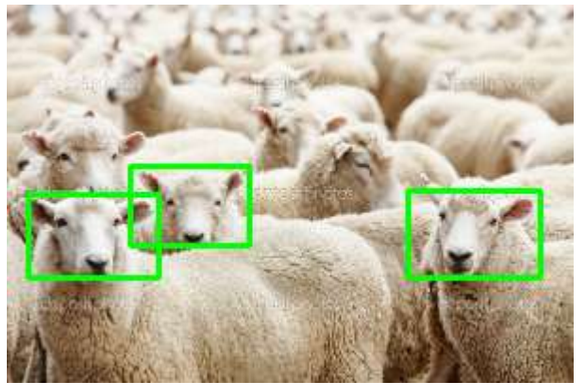
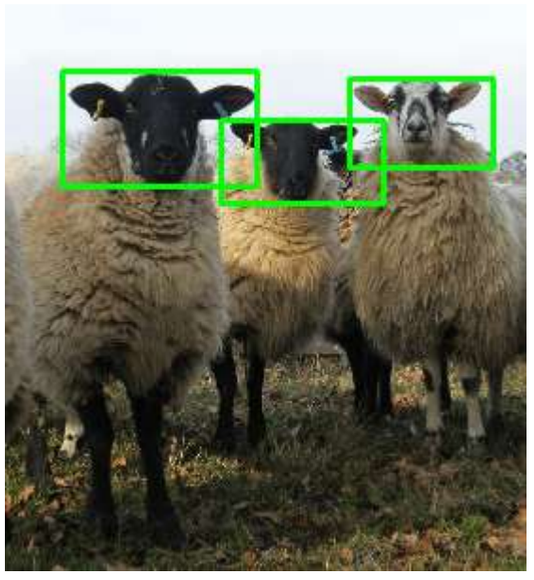
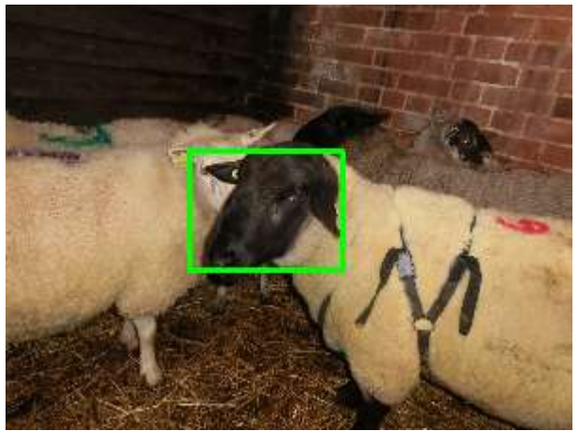
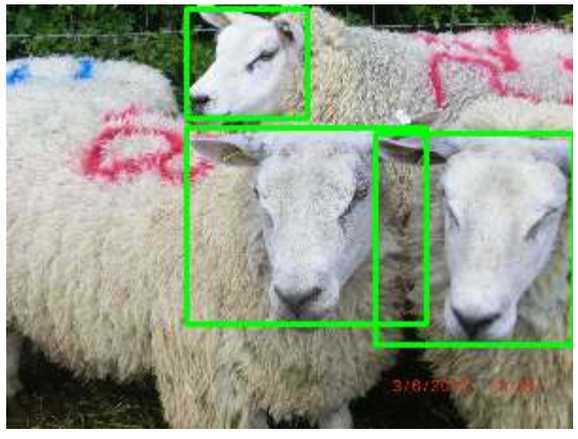




[Yiting Lu, Marwa Mahmoud & Peter Robinson. ‘Estimating sheep pain level using facial action unit detection’, FG 2017]

Animal face detection

- Small dataset to train an end-to-end Neural Network
- Fine-tuning a convolutional neural network model trained on a large human face detection dataset
- Single Shot MultiBox Detector (SSD) built on the MobiNet architecture – Fast and Accurate

Facial landmarks & head pose detection

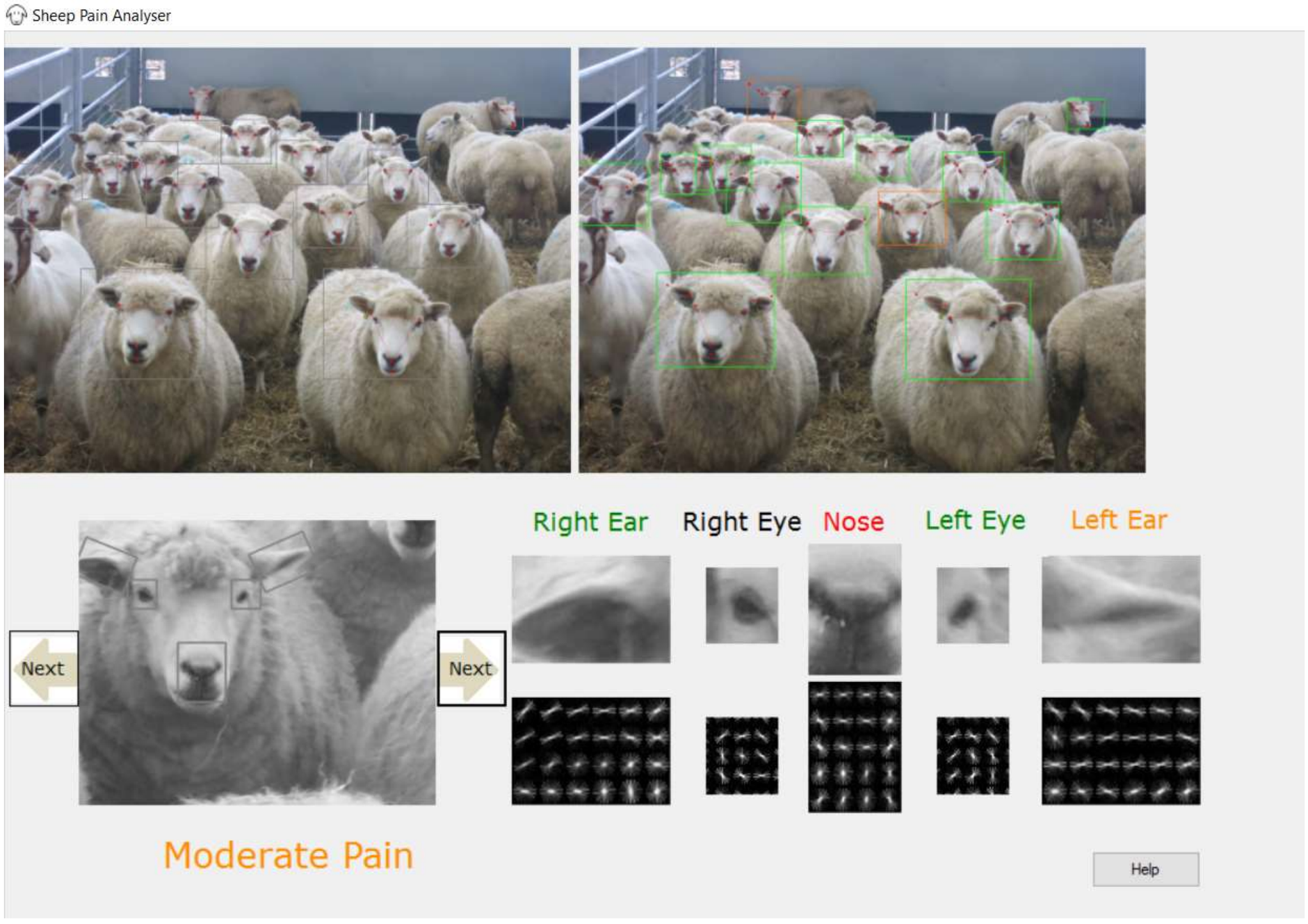
Charlie Hewitt and Marwa Mahmoud. Pose-Informed Face Alignment for Extreme Head Pose Variations in Animals. ACII 2019.

Facial expressions of pain in sheep

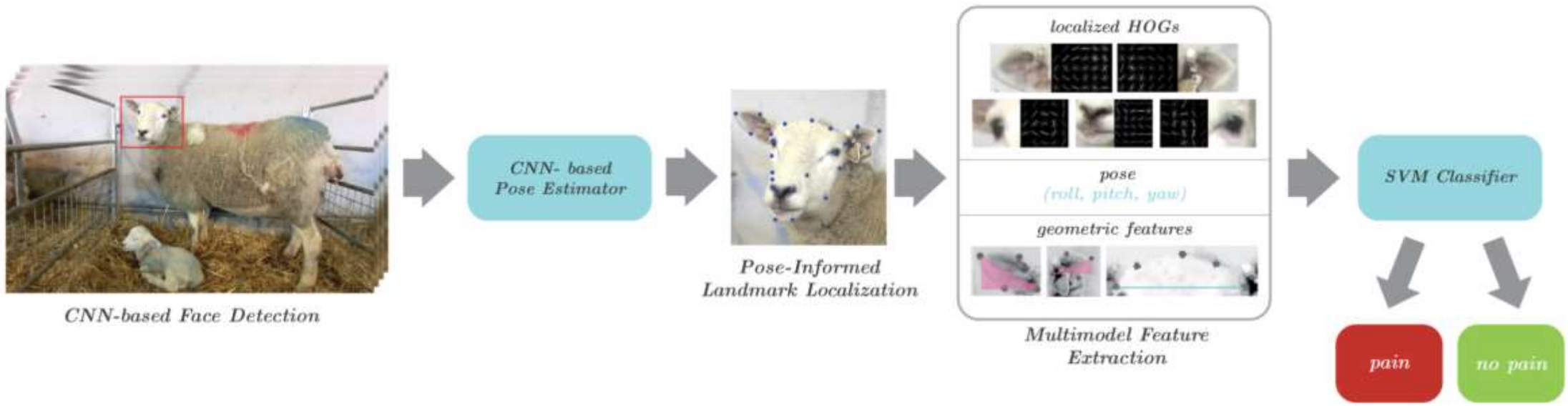
In pain ☹ *Happy ☺*

In pain ☹ *Happy ☺*

Pipeline evaluation in the wild – BBC Countryfile

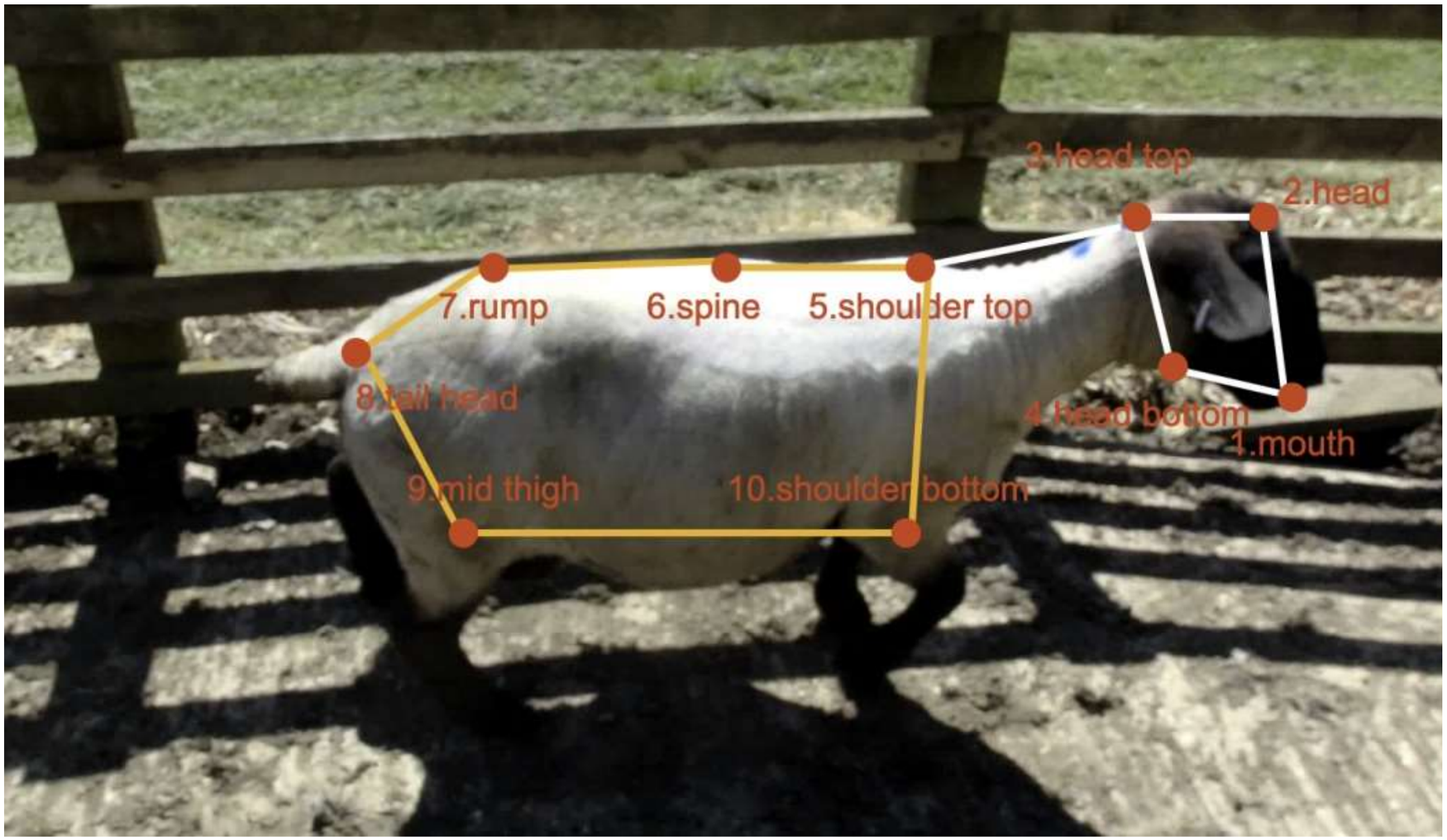


Towards automatic monitoring of disease progression in sheep



Francisca Pessanha, Krista McLennan and Marwa Mahmoud. Towards automatic monitoring of disease progression in sheep: A hierarchical model for sheep facial expressions analysis from video. FG 2020.

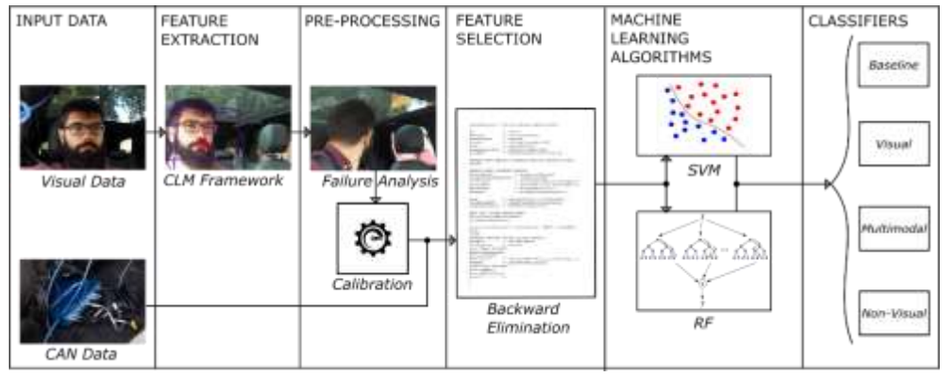
Body expressions/ multimodal analysis



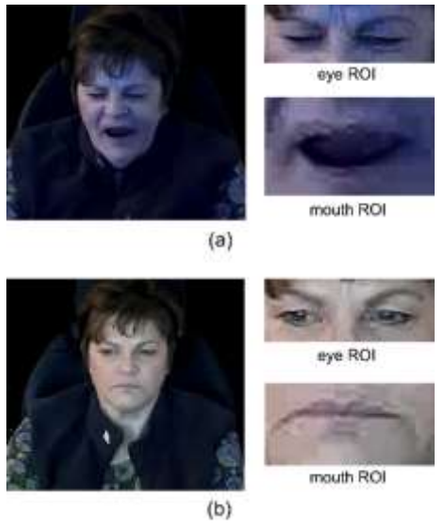
Enhancing Driver's Experience using vision research

Computer Vision for Automotive

Monitoring Driver's Distraction



Drowsiness Detection



Acknowledgements

@marwammmahmoud

Francesco Ceccarelli

Weizhe Lin

Indigo Orton

Francisca Pessanha

Bruno Tarfur

Prof Mark Johnson
Dept. of Psychology
University of Cambridge

Dr Krista McLennan
Dept. of Biological Sciences
University of Chester

Dr Gabriela Pavarini
Dept. of Psychiatry
University of Oxford

Dr Staci M. Weiss
Dept. of Psychology
University of Cambridge