



# Continuously Reproducing Toolchains in Pattern Recognition and Machine Learning Experiments

André Anjos, Manuel Günther, Tiago de Freitas Pereira,  
Pavel Korshunov, Amir Mohammadi, and Sébastien Marcel  
Idiap Research Institute, Switzerland



Start here: <https://www.idiap.ch/software/bob/>

## Typical Paper

### A Scalable Formulation of Probabilistic Linear Discriminant Analysis: Applied to Face Recognition

Laurent El Shafey, Chris McCool, Roy Wallace, and Sébastien Marcel

**APPENDIX A  
MATHEMATICAL DERIVATIONS**  
The goal of the following section is to provide more detailed proofs of the formulae given in the article for both training and computing the likelihood.

The following proofs make use of a formulation of the inverse of a block matrix that uses the Schur complement. The corresponding identity can be found in [1] (Equations 1.11 and 1.10).

where we have substituted  $R = (L + MO^T N)^{-1}$ . Another related expression is the Woodbury matrix identity (Equation C.7 of [2]), which states that,

$$(L + MON)^{-1} = L^{-1} - L^{-1}M(O^{-1} + NL^{-1}M)^{-1}NL^{-1}, \quad (52)$$

**A. Scalable training**  
The bottleneck of the training procedure is the expectation step (E-Step) of the Expectation-Maximization algorithm. This E-Step requires the computation of the first and second order moments of the latent variables.

1) *Estimating the first order moment of the Latent Variables:* The most computationally expensive part when estimating the latent variables is the inversion of the matrix  $\tilde{P}$  (Equation (27)). This matrix is block diagonal, the two blocks being  $\tilde{P}_0$  (Equation (28)) and (a repetition of)  $\tilde{P}_1$  (Equation (29)).

$$\tilde{P} = \begin{bmatrix} \tilde{P}_0 & 0 & \dots & 0 \\ 0 & \tilde{P}_1 & \ddots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \tilde{P}_1 \end{bmatrix}, \quad (53)$$

The inverse of  $\tilde{P}_0$  is equal to the matrix  $G$ , defined by (30). This matrix is of constant size ( $D_0 \times D_0$ ), irrespective

of the number of training samples for the class. In addition, the inversion of  $\tilde{P}_0$  can be further optimised using the block matrix inversion identity introduced at the beginning of this section, leading to

$$\tilde{P}_0^{-1} = \begin{bmatrix} \mathcal{F}_0 & \sqrt{J}\mathcal{H}^T \\ \sqrt{J}\mathcal{H} & (I_{D_0} - J_0\mathcal{H}\mathcal{F}_0^T\Sigma^{-1}G)\mathcal{G} \end{bmatrix}, \quad (54)$$

where  $\mathcal{F}_0$  is defined by (33) and  $\mathcal{H}$  by (37). Then, the computation of  $\tilde{P}^{-1}\tilde{A}^T\Sigma^{-1}$  gives a block diagonal matrix, the first block being

$$\begin{bmatrix} \sqrt{J}\mathcal{F}_0\mathcal{F}_0^T\Sigma^{-1} & \sqrt{J}\mathcal{H}\mathcal{F}_0^T\Sigma^{-1} \\ \sqrt{J}\mathcal{H}\mathcal{F}_0^T\Sigma^{-1} & (I_{D_0} - J_0\mathcal{H}\mathcal{F}_0^T\Sigma^{-1}G)\mathcal{G} \end{bmatrix},$$

and the other ones being equal to  $\mathcal{G}\mathcal{G}^T\Sigma^{-1}$ . As explained in section II.B.a of the article,  $h_1$  corresponds to the upper sub-vector of  $\tilde{y}_1$  and is not affected by the change of variable, as depicted in (21). Therefore, the first order moment of  $h_1$  is directly obtained by multiplying the first block-rows of the matrix  $\tilde{P}^{-1}\tilde{A}^T\Sigma^{-1}$  with  $\tilde{a}_1$ , which gives (31).

Considering only the  $\tilde{a}_1$  (lower) sub-vector of  $\tilde{y}_1$ , the corresponding (lower) part  $\tilde{B}$  of the matrix  $\tilde{P}^{-1}\tilde{A}^T\Sigma^{-1}$  can be decomposed into a sum of two matrices, the first one being sparse with a single non-zero block (upper left) equal to  $\tilde{B}_0 = -\sqrt{J}\mathcal{G}\mathcal{G}^T\Sigma^{-1}\mathcal{F}_0\mathcal{F}_0^T\Sigma^{-1}$ , and the second one being diagonal by blocks with identical blocks  $\tilde{B}_1 = \mathcal{G}\mathcal{G}^T\Sigma^{-1}$ .

$$\tilde{B} = \begin{bmatrix} \tilde{B}_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} \tilde{B}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tilde{B}_1 \end{bmatrix}. \quad (55)$$

Furthermore, the first order moment of the variables  $w_i$  is given by

$$E[\tilde{w}_i|\tilde{a}_1, \Theta] = (\tilde{U}^T \otimes I_{D_0}) \begin{bmatrix} \tilde{B}_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tilde{a}_1 \quad (56)$$

$$+ (\tilde{U}^T \otimes I_{D_0}) \begin{bmatrix} \tilde{B}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tilde{B}_1 \end{bmatrix} (\tilde{U} \otimes I_{D_0}) \tilde{a}_1.$$

The previous decomposition greatly simplifies the computation, and leads to the following expression for each  $w_{i,j}$ .

$$E[w_{i,j}|\tilde{a}_1, \Theta] = -\mathcal{G}\mathcal{G}^T\Sigma^{-1}\tilde{a}_{1,j} - \mathcal{G}\mathcal{G}^T\Sigma^{-1}\mathcal{F}_0\mathcal{F}_0^T\Sigma^{-1}\tilde{a}_{1,j} \quad (57)$$

JOURNAL TO BE REFINED, VOL. X, NO. 1, JANUARY 2013

12

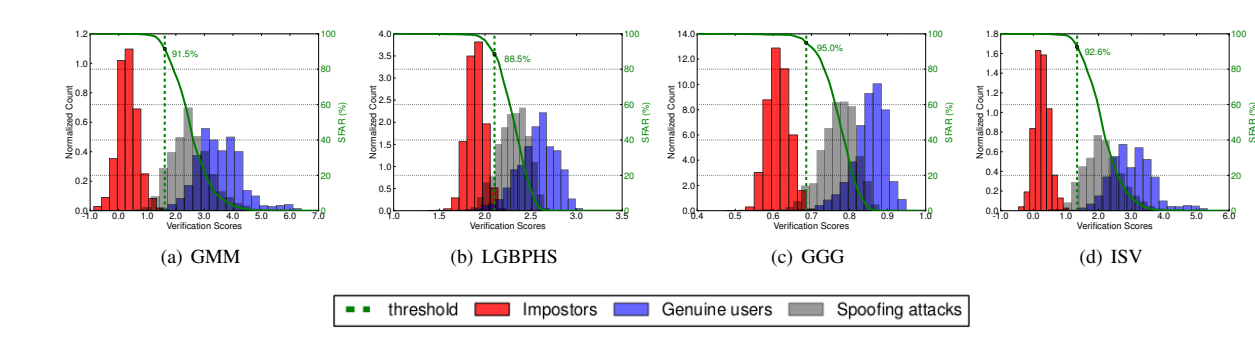


Fig. 10: Score distributions of baseline face verification systems. The full green line shows how SFAR changes with moving the threshold.

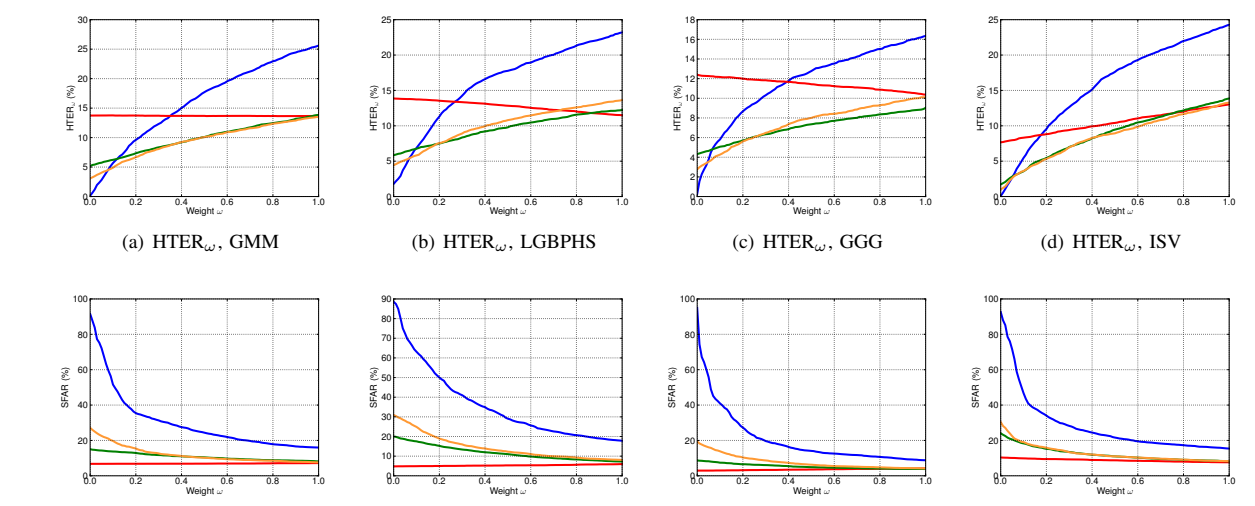


Fig. 12: EPSC for comparison of fusion techniques of baselines with LBP anti-spoofing algorithm

#### D. Performance of fused systems

In our last experiment, we compare the four face verification systems when fused with ALL counter-measures using PLR fusion scheme. Firstly, we illustrate how fusion changes the score distribution for each of them separately in Figure 14. Then, in Figure 15 we compare which of the fused systems performs the best.

While Figure 10 shows that the spoofing attacks of Replay-Attack are in the optimal category when fed to the baseline face verification systems, Figure 14 illustrates that their effectiveness has vastly changed after fusion. The score distribution of the spoofing attacks is now mostly overlapping with the score distribution of the zero-effort impostors, allowing for better discriminability between the positive class and the two negative classes. The results are reflecting this observation:

while HTER<sub>0</sub> increases rapidly with  $\omega$  and reaches up to 25% for some of the baseline systems, it increases very mildly and does not exceed 4.1% for the fused systems. The major augmentation of the robustness to spoofing of the systems after

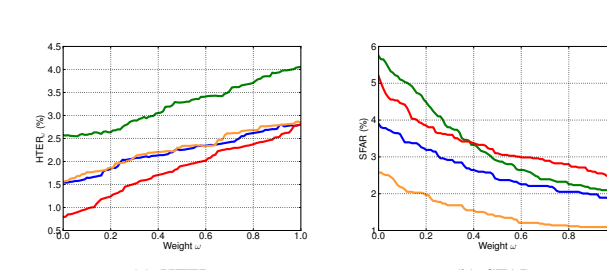


Fig. 15: EPSC curves to compare fused systems

- Pattern recognition and machine learning research work often contains experimental results on real-world data, which corroborates hypotheses and provides a canvas for the development and comparison of new ideas.
- Reproducibility is often overlooked.
- Scientific experiments often consist of many steps and parameters that are difficult to report.
- In our view, **reproducible research work should be repeatable, shareable, extensible and stable**.
- We investigate the implications of this key properties of RR work and the requirements for frameworks to fulfil them.

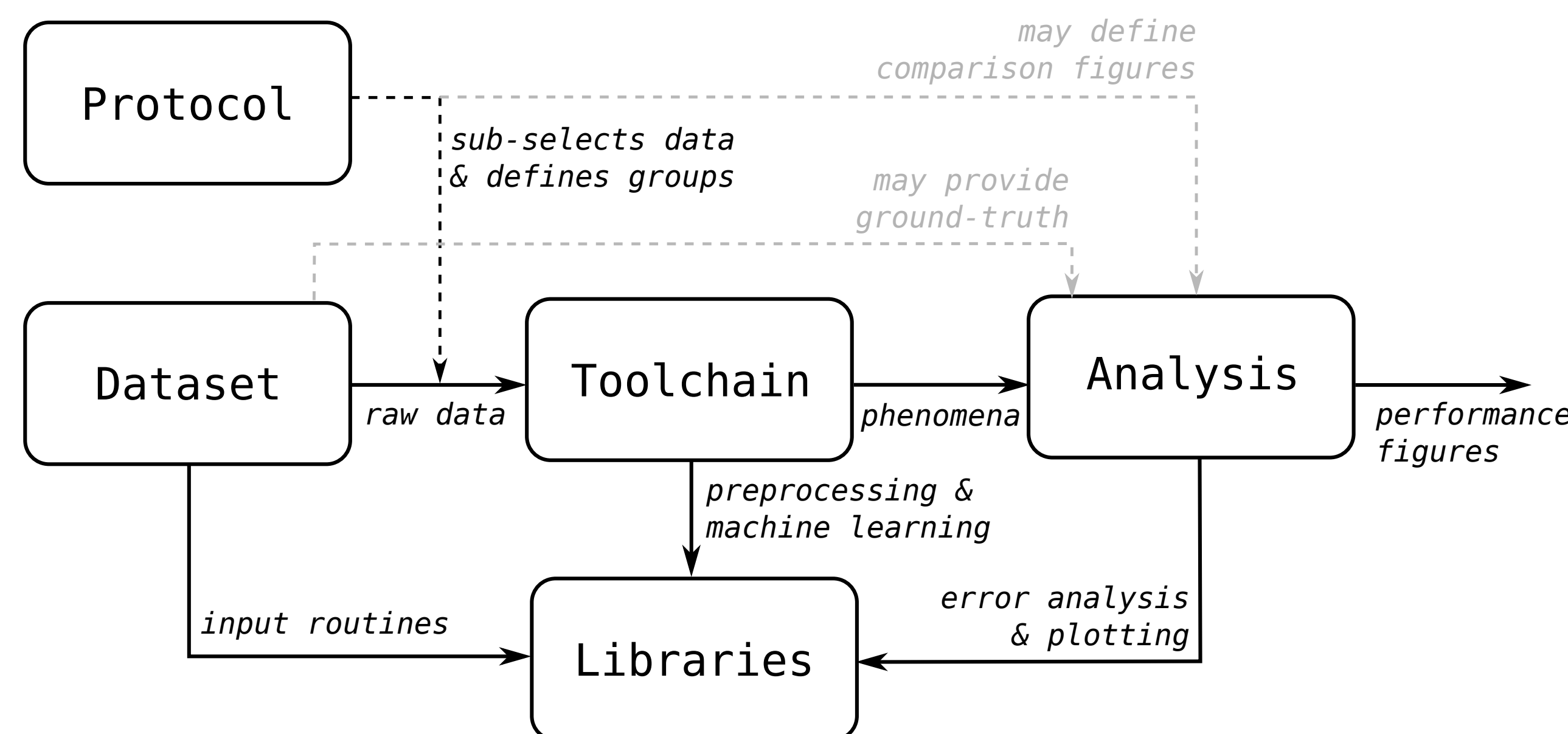
## Definition

We consider a published paper to be reproducible if it is:

- Repeatable:** It should be **possible to re-run** experiments declared in a report and obtain the same results, given the same selection of data, software, hyper-parameters, and evaluation protocol.
- Shareable:** It should be **possible to share** the material (**data and code**) with others, so they can repeat experiments declared in a report.
- Extensible:** It should be **easy to implement new research directions or evaluate new conditions** on existing experimental infrastructure.
- Stable:** The **repeatability through time should be guaranteed**, on a best-effort basis.

## Framework

Extensibility is an important aspect of reproducibility as *today's state of the art will eventually become tomorrow's baselines*. A "workflow" needs to be in place:



## Tools

Bare software frameworks are seldom successful in the long run without committed people and accompanying tools:

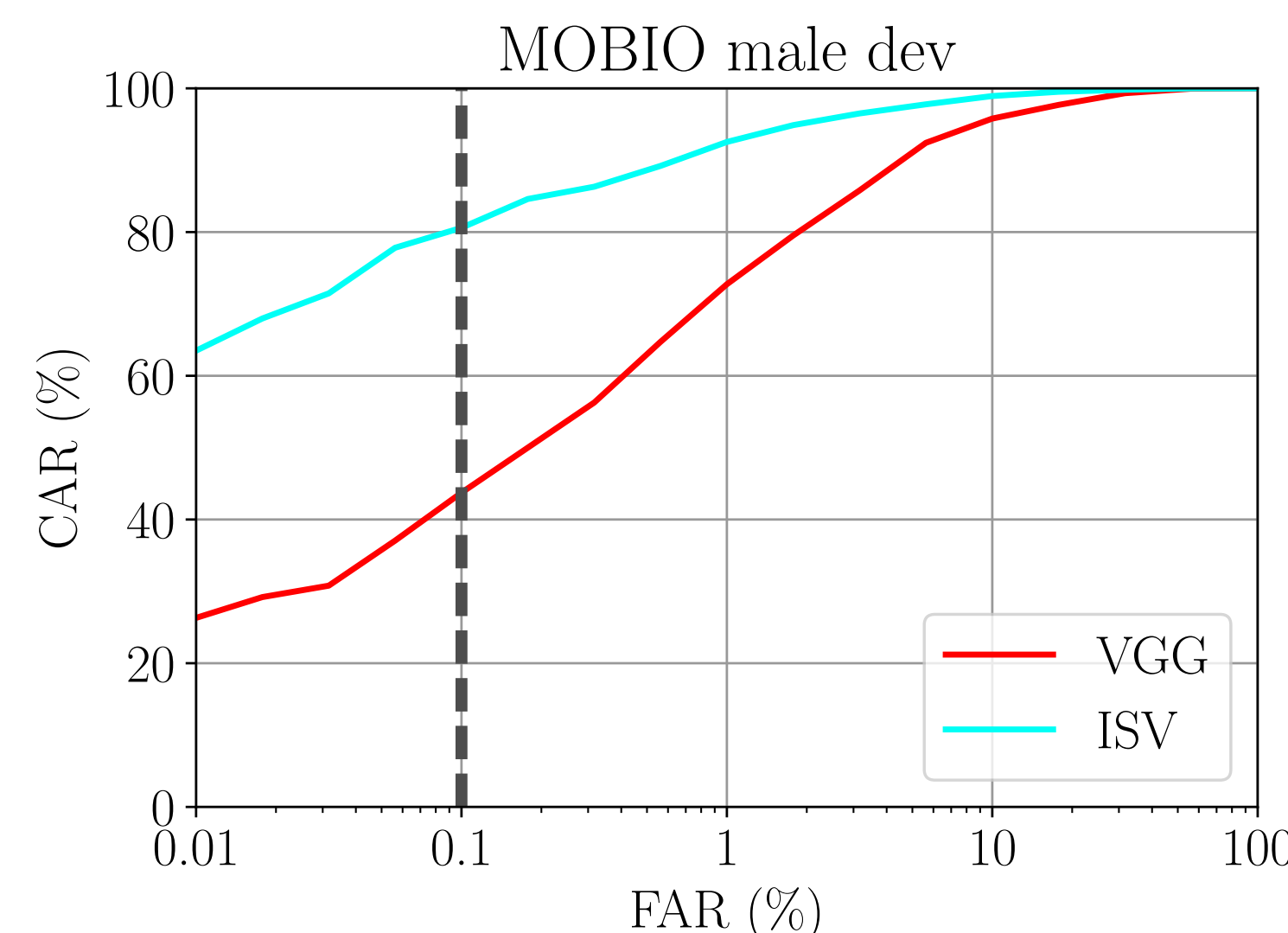
- Version control
- Unit testing and quality control
- Packaging and deployment
- Documentation

## Use-case Analysis: Reproducible face recognition experiments

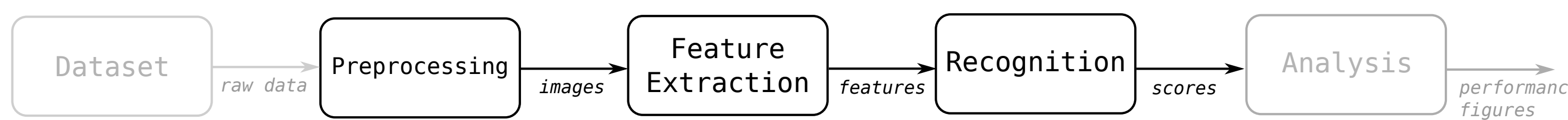
- Get data: <https://www.idiap.ch/dataset/mobio>
- Get code: <https://gitlab.idiap.ch/bob/bob.paper.icml2017>

- Install: Trivially done on any Linux/Mac OSX (64-bits) with Conda
- Run: Single command line to launch the multi-stage pipeline. **Automatic parallelization** on SGE or local machine supported

### Results:



### Toolchain implemented by framework:



### Version and Quality control:

## Conclusions

- To guarantee long-time reproducibility, scientific **work must be repeatable, shareable, extensible and stable over time**.
- Committing to continuous reproducibility implies the creation and maintenance of a **re-usable framework**.
- Efforts can be reduced if **researchers are unified** into standardized frameworks.
- Bob** is an *example*; we expect others to appear in the future.
- Web-based frameworks such as the **BEAT** platform may provide a simpler mechanism of reproducibility while implementing "social" research and development.