How Artificial Intelligence can be used for Behavioral Identification?

Yris Brice Wandji Piugie[1,2], Joël Di Manno[2] , Christophe Rosenberger[1] and Christophe Charrier[1]

[1]Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, FRANCE

[2]FIME SAS, Caen, FRANCE

{brice.wandji, joel.dimanno}@fime.com, christophe.rosenberger@ensicaen.fr, christophe.charrier@unicaen.fr

GREYC
Electronics and Computer Science Laboratory

Overall Goal : Identify a User knowing her/his Behavior

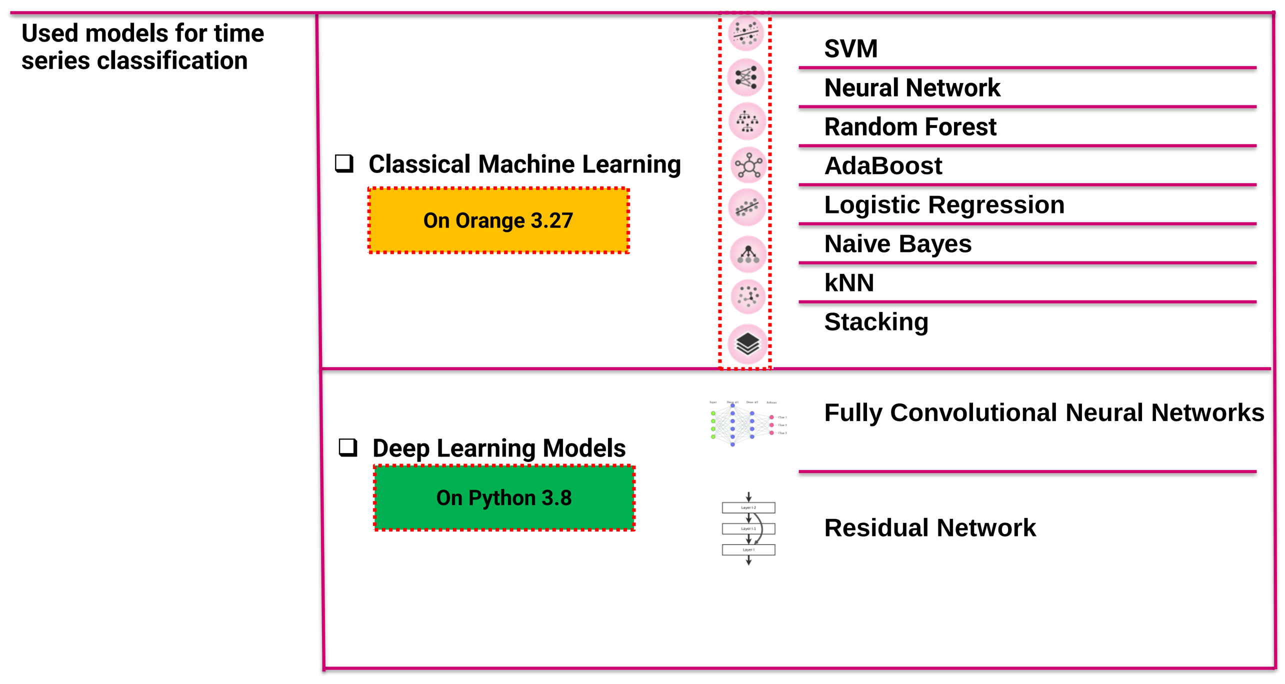
1. Context and Problematic

❑ **Behavioral biometrics**

❑ **Problematic**

User identification considering their behaviors

2. Comparative Study



3. Generic Workflow for ML

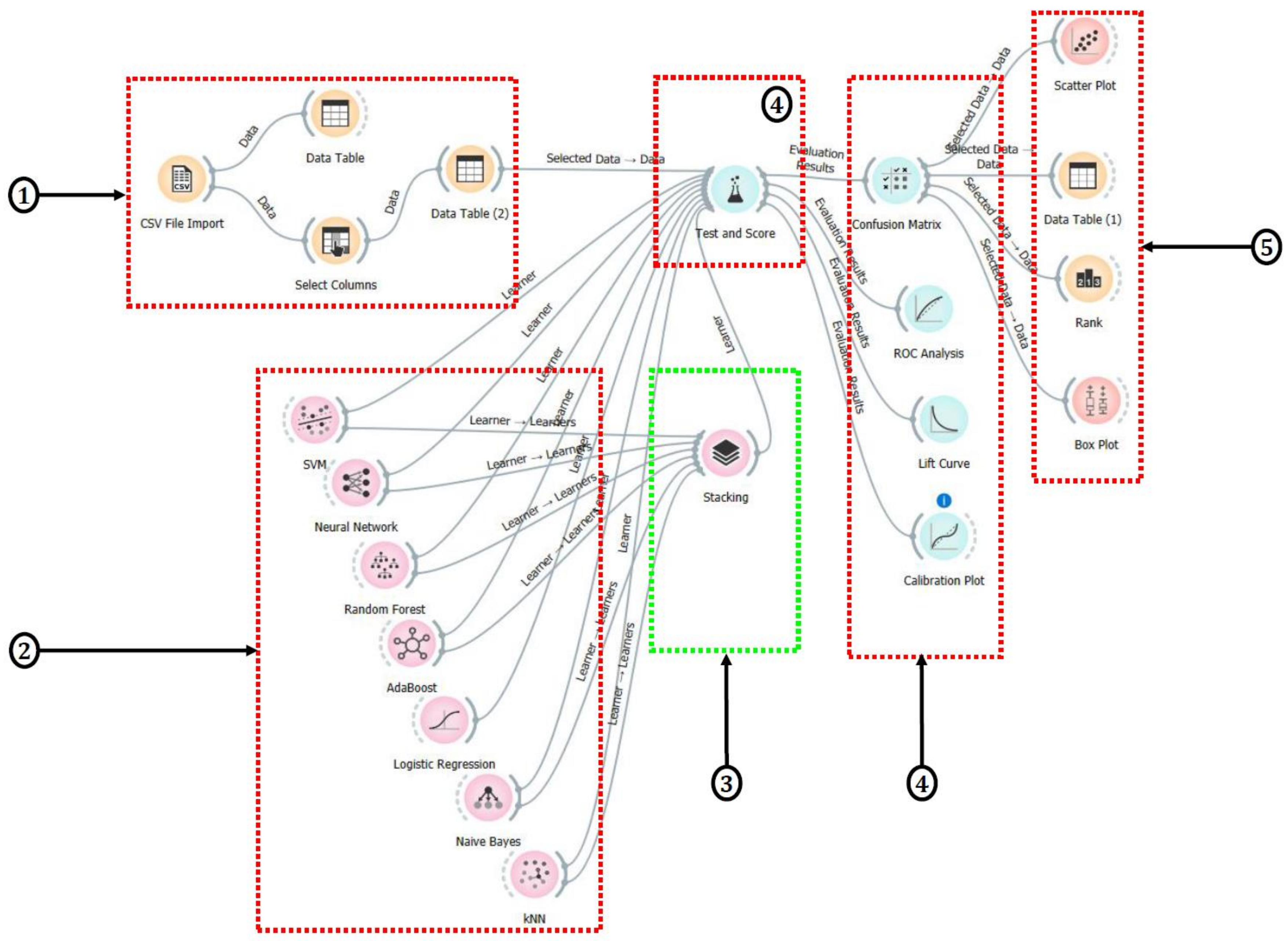
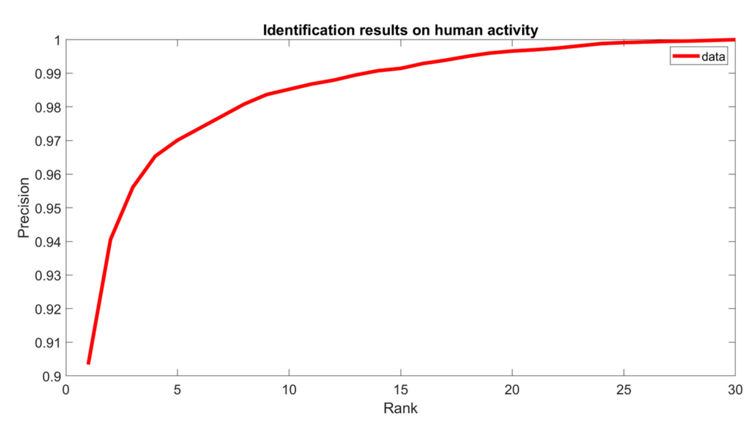


Figure 1. Global workflow for user identification from behavioral data

7. CMC curve on behavioral biometrics data

❑ **HAR database**



❑ **GREYC-NISLAB database**

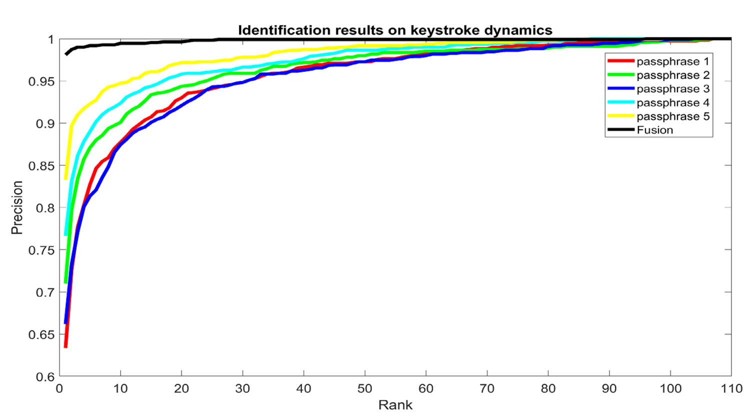
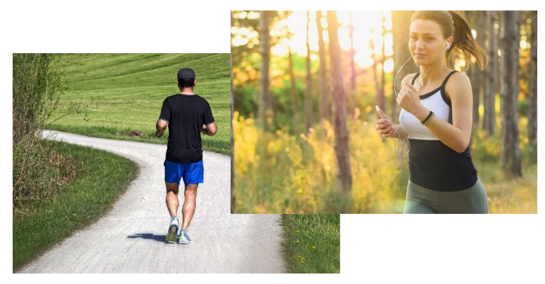


Fig 4. CMC curve of Stacking model in Orange workflow

4. Protocol description

❑ Training set : 70% per user data in the dataset
❑ Testing set : 30% per user data in the dataset

❑ **Human activities – HAR database**



Activities
Laying
Sitting
Standing
Walking
Walking Downstairs
Walking Upstairs

30 users

❑ **Keystroke dynamics – GREYC-NISLAB database**

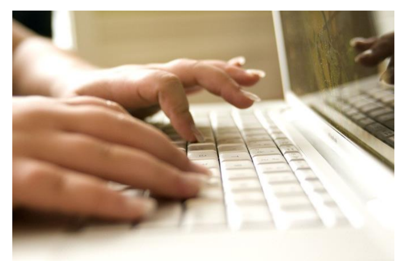


Table III: Passphrases

Password	Description	Size	Features
P1	leonardo dicaprio	17-char	64
P2	the rolling stones	18-char	68
P3	michael schumacher	18-char	68
P4	red hot chilli peppers	22-char	84
P5	united states of america	24-char	92
P_T	fusion of features (P1+P2+P3+P4+P5)	99-char	376

110 users

5. Performance Metrics

❑Classification Accuracy (A or CA)

$$A = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (1)$$

❑Precision score (P)

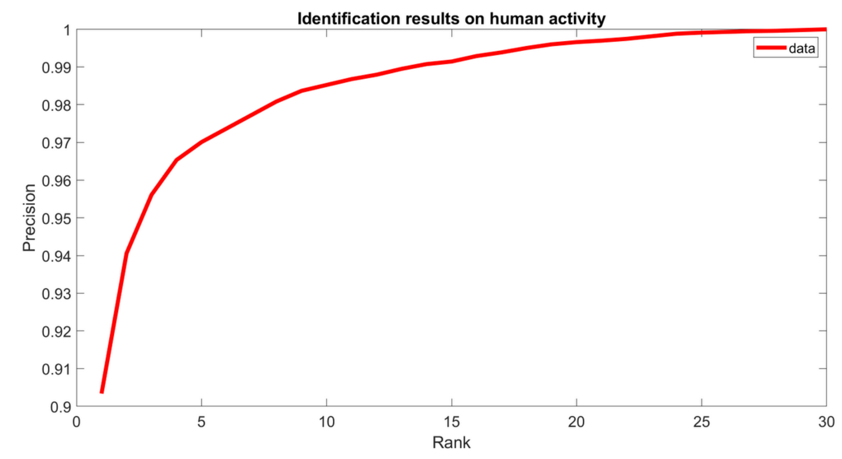
$$P = \frac{T_P}{T_P + F_P} \quad (2)$$

❑Recall (R)

$$R = \frac{T_P}{T_P + F_N} \quad (3)$$

❑Area Under the Curve (AUC)

❑Cumulative Match Characteristic (CMC) Curve



6. Experimental Results

❑**Deep Learning Models**

Table X: UCI-HAR and GREYC-NISLAB deep performance metrics

Dataset	Classifier name	CA (%)	P (%)	R (%)
HAR	ResNet	87.05	87.20	86.73
	FCN	68.58	80.09	68.24
P_T	ResNet	80.30	82.94	82.23
	FCN	76.06	78.95	79.01

Deep Learning Methods give good results but not exceptional !

❑ **Classical Machine Learning**

- **HAR database**

Table VII: User identification performance metrics with Orange workflow on HAR dataset from human activities.

Model	AUC (%)	CA (%)	P (%)	R (%)
Stack	99.65	93.89	93.90	93.89
Neural Networks	98.97	87.75	87.73	87.75
Random Forest	98.21	85.78	85.89	85.78
kNN	97.56	81.40	82.07	81.40
AdaBoost	89.33	81.06	81.25	81.06
SVM	96.57	78.45	80.22	78.45
Logistic Regression	96.76	78.38	78.40	78.38
Naive Bayes	79.24	41.33	48.49	41.33

By analyzing users activities, and merging all the models, in 93.90% of the cases we can recognize a person among the 30 users.

- **GREYC-NISLAB database**

Table VIII: User identification performance metrics with Orange workflow on GREYC-NISLAB from keystroke dynamics.

Password database	Model	AUC (%)	CA (%)	P (%)	R (%)
P1	Stack	96.22	63.09	63.67	63.10
P2	Stack	99.08	69.73	72.15	69.73
P3	Stack	98.49	63.91	66.10	63.91
P4	Stack	99.22	77.73	79.64	77.73
P5	Stack	98.56	83.73	84.30	83.73
P_T	Stack	99.99	98.10	98.3	98.10

We are able to identify one user among the 110 users with a goal rate over the 98% of the cases.

Conclusions

Targets	CA(%)
Subject	98.18
Handedness	99.27
Gender	88.73
Age	70.73

Traditional machine learning tools can have a significant impact on a person's privacy!

Normandie Université

PIUGIE, Yris Brice Wandji, DI MANNO, Joël, ROSENBERGER, Christophe, et al. How Artificial Intelligence can be used for Behavioral Identification?. In : 2021 International Conference on Cyberworlds (CW). IEEE, 2021. p. 246-253.