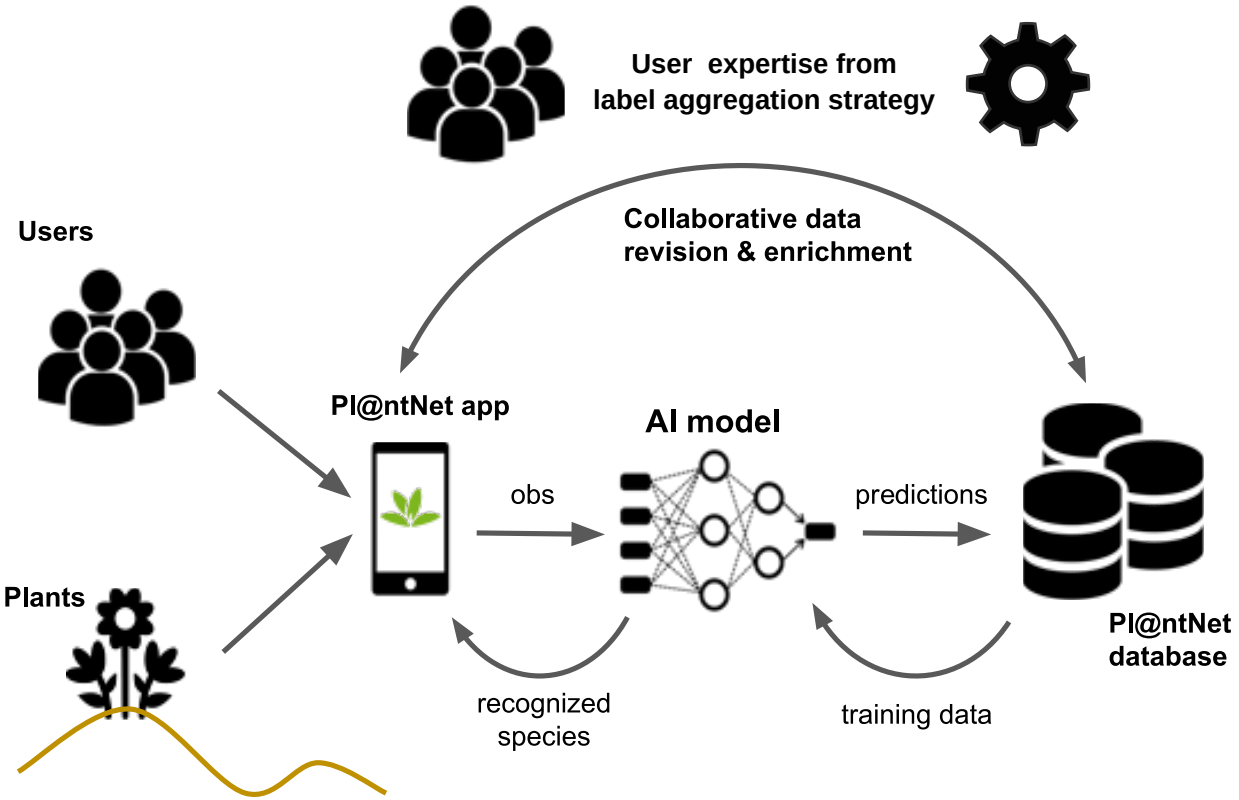
Cooperative data enrichment algorithms

Tanguy Lefort
IMAG, Univ Montpellier, CNRS
Inria, LIRMM,

Ongoing work with...

- Alexis Joly
- Benjamin Charlier
- Joseph Salmon
- Pierre Bonnet
- Antoine Affouard
- JC Lombardo

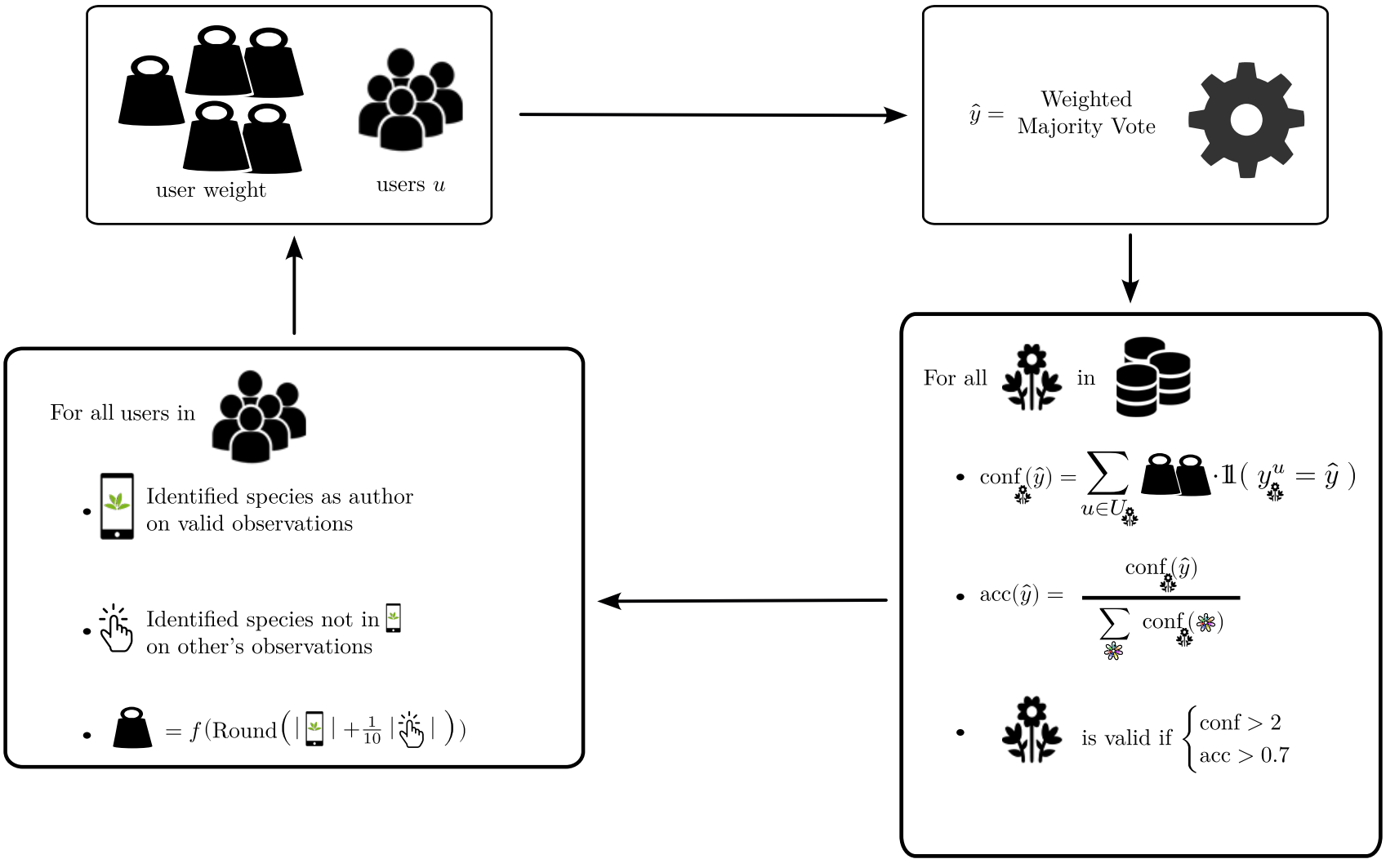
Key concept of Pl@ntNet: Collaborative AI



Pl@ntNet label aggregation

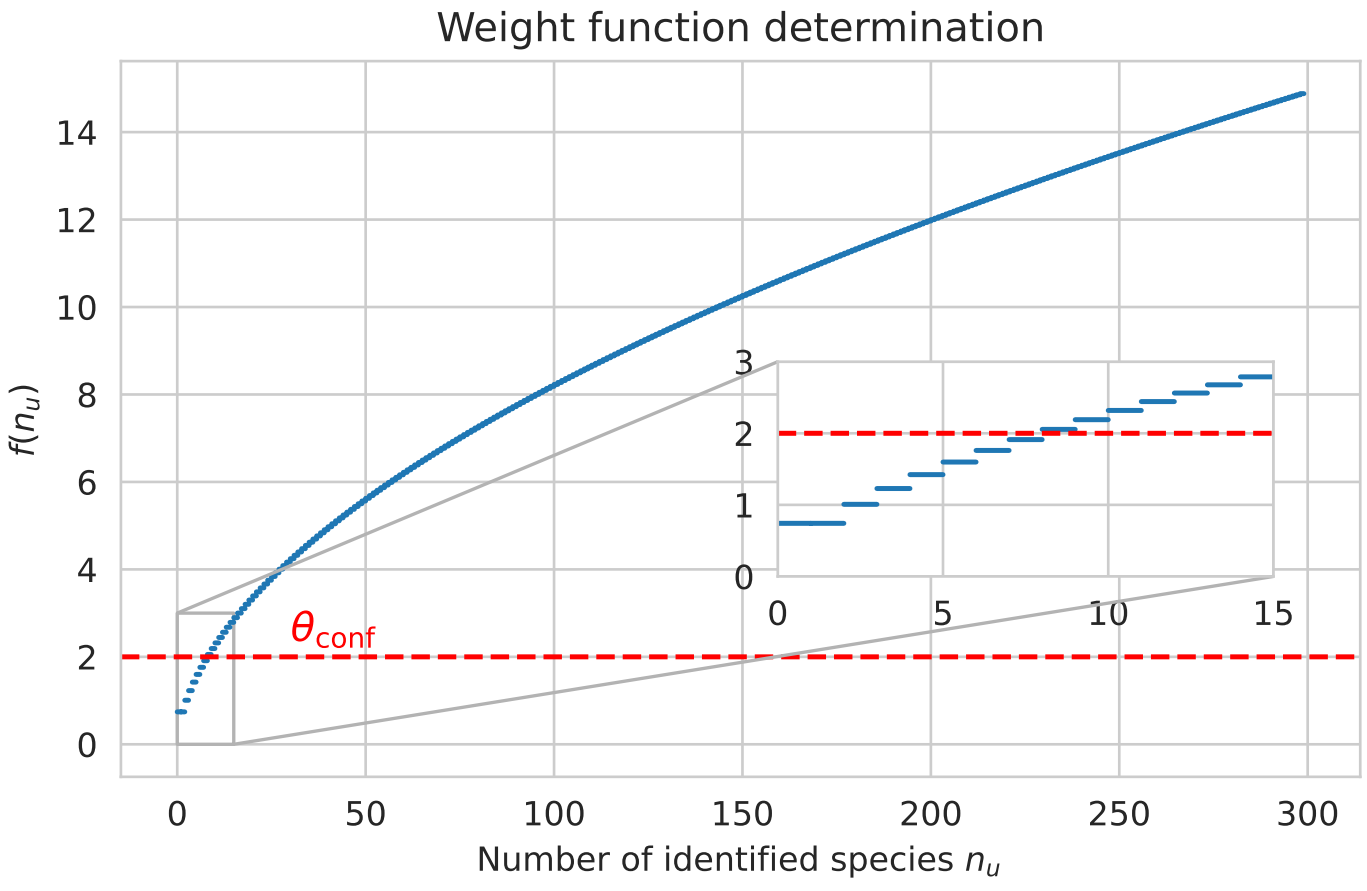
EM based algorithm

Weighting users vote by their estimated number of identified species



Choice of weight function

$$f(n_u) = n_u^{\alpha} - n_u^{\beta} + \gamma \text{ with } \begin{cases} \alpha = 0.5 \\ \beta = 0.2 \\ \gamma = \log(1.7) \simeq 0.74 \end{cases}$$



Other existing strategies

- **Majority Vote** (MV)

Other existing strategies

- **Majority Vote** (MV)
- **Worker agreement with aggregate (WAWA)** (Appen 2021)
  - Majority vote
  - Weight user by how much they agree with the majority
  - Weighted majority vote

Other existing strategies

- **Majority Vote** (MV)
- **Worker agreement with aggregate (WAWA)** (Appen 2021)
  - Majority vote
  - Weight user by how much they agree with the majority
  - Weighted majority vote
- **TwoThird** (cornerstone of iNaturalist)
  - Need 2 votes
  - 2/3 of agreements

Extracting a subset of a Pl@ntNet

Design and some numbers

- South Western European flora obs since 2017
- 823 000 users answered more than 11000 species
- 6 700 000 observations
- 9 000 000 votes casted
- **Imbalance**: 80% of observations are represented by 10% of total votes

Extracting a subset of a Pl@ntNet

Design and some numbers

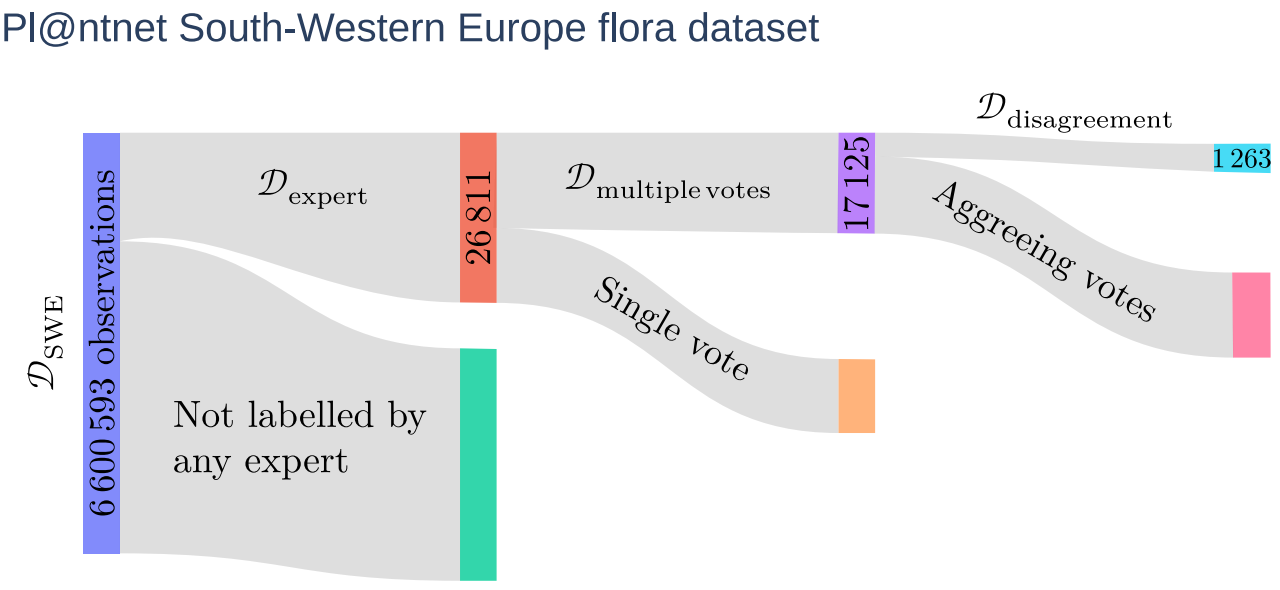
- South Western European flora obs since 2017
- 823 000 users answered more than 11000 species
- 6 700 000 observations
- 9 000 000 votes casted
- **Imbalance**: 80% of observations are represented by 10% of total votes

No ground truth available to evaluate the strategies

Extracting a subset of a Pl@ntNet

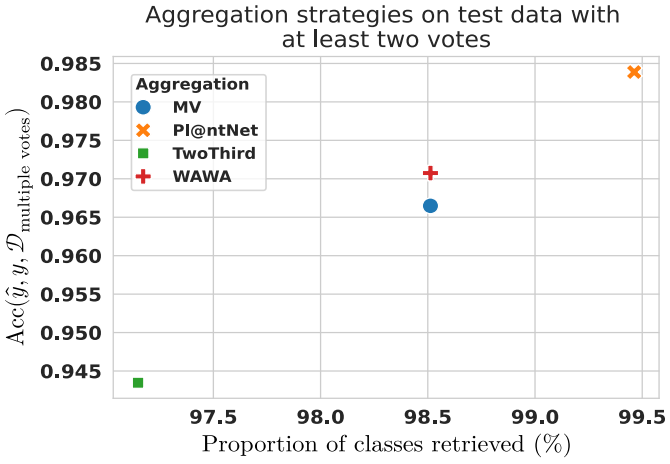
Creation of test sets

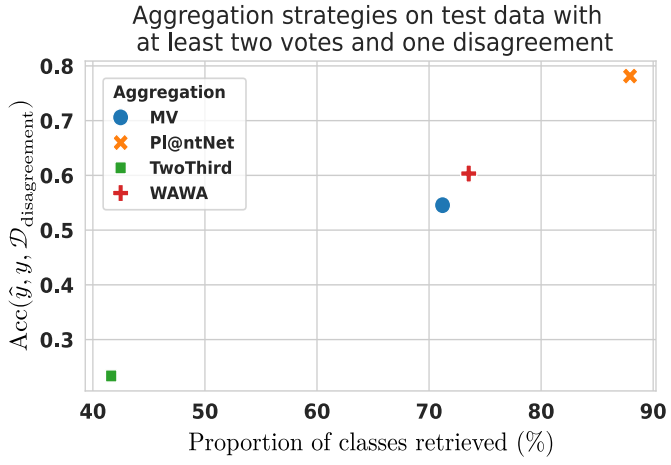
- Extraction of 98 experts (TelaBotanic + prior knowledge – thanks to Pierre Bonnet)



Performance

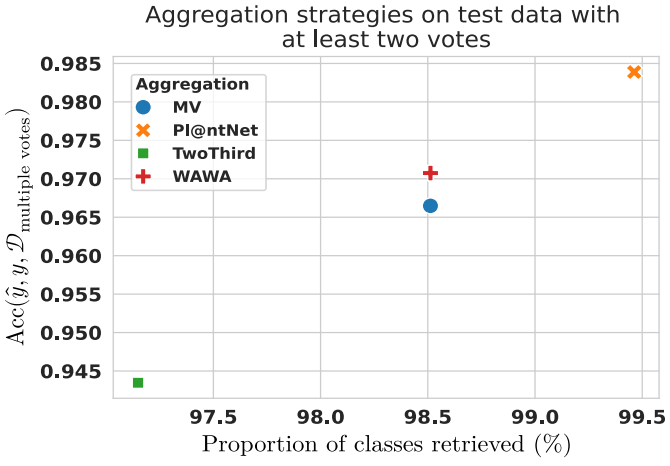
Accuracy and volume of classes kept

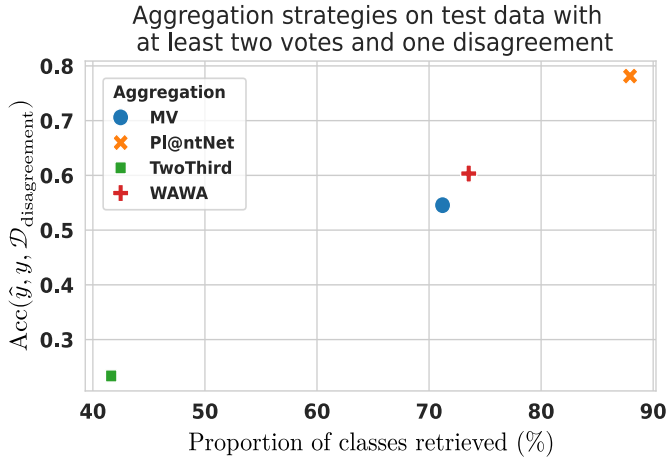




Performance

Accuracy and volume of classes kept



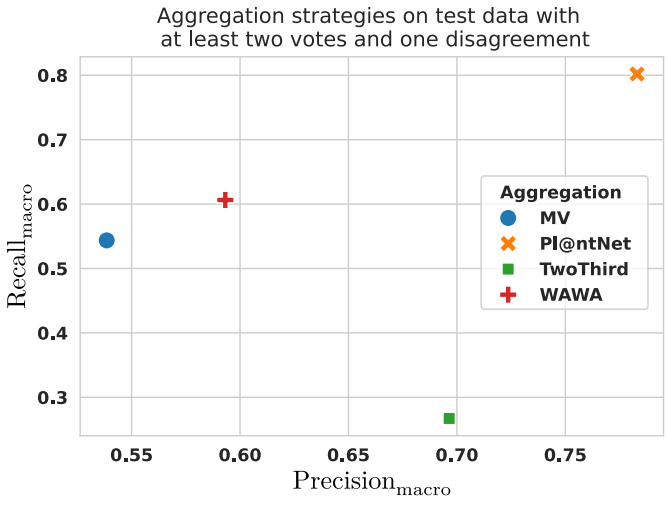


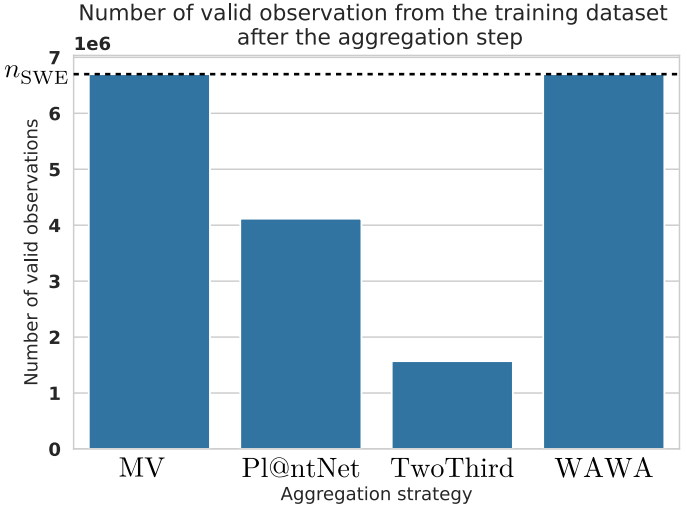
In short

- Pl@ntNet aggregation performs better overall
- TwoThird is highly impacted by their reject threshold
- In ambiguous settings (right), strategies weighting users are better

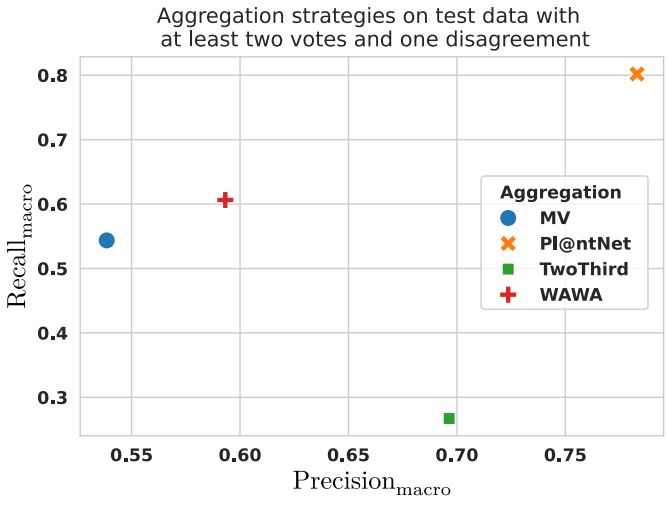
Performance

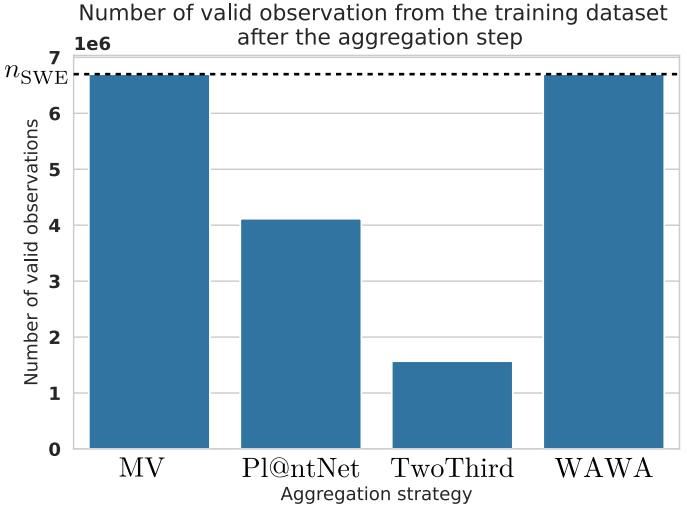
Precision, recall and validity





Performance
Precision, recall and validity





In short

- Pl@ntNet aggregation performs better overall
- TwoThird has good precision but bad recall
- We indeed remove some data but less than TwoThird

Thanks