

Improving Nearest Neighbor Classification using Ensembles of Evolutionary Generated Prototype Subsets

Sarah Vluymans^{1,2,3}, Nele Verbiest¹, Chris Cornelis^{1,3}, Nicolás García-Pedrajas⁴ and Yvan Saeys^{2,5}

¹ Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium

² Data mining and Modeling for Biomedicine, VIB Inflammation Research Center, Belgium

³ Department of Computer Science and Artificial Intelligence, University of Granada, Spain

⁴ Department of Computer Science and Numerical Analysis, University of Córdoba, Spain

⁵ Department of Internal Medicine, Ghent University, Belgium

Contact : sarah.vluymans@ugent.be



Prototype selection (PS)

- Preprocessing step: select a subset of useful learning instances
- Remove noise and redundancy
- Commonly used with nearest-neighbor classification

State-of-the-art: genetic PS algorithms

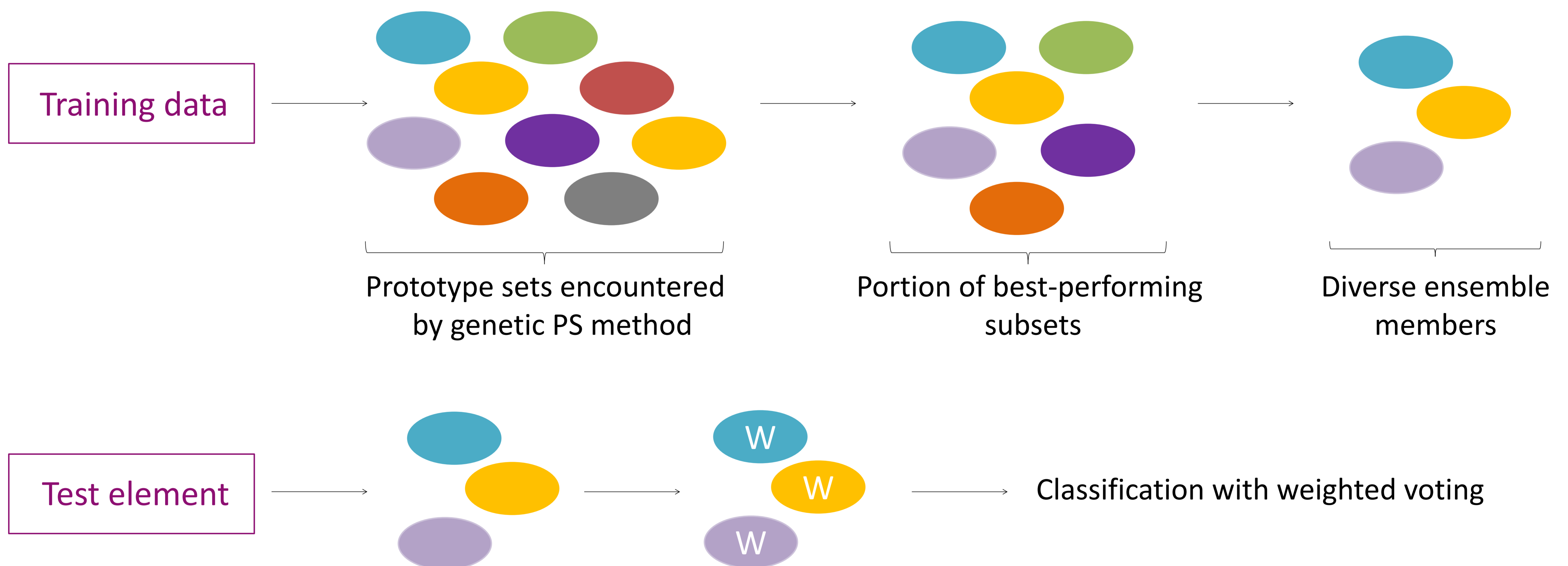
High reduction + high prediction performance of subsequent classifier

Drawback

Many candidate subsets are constructed, but only one is used in the end

Subsets that are not globally optimal could still perform well in certain regions
 → use a diverse ensemble of well-performing prototype subsets

Ensemble of Evolutionary Generated Prototype Subsets (EEGPS)



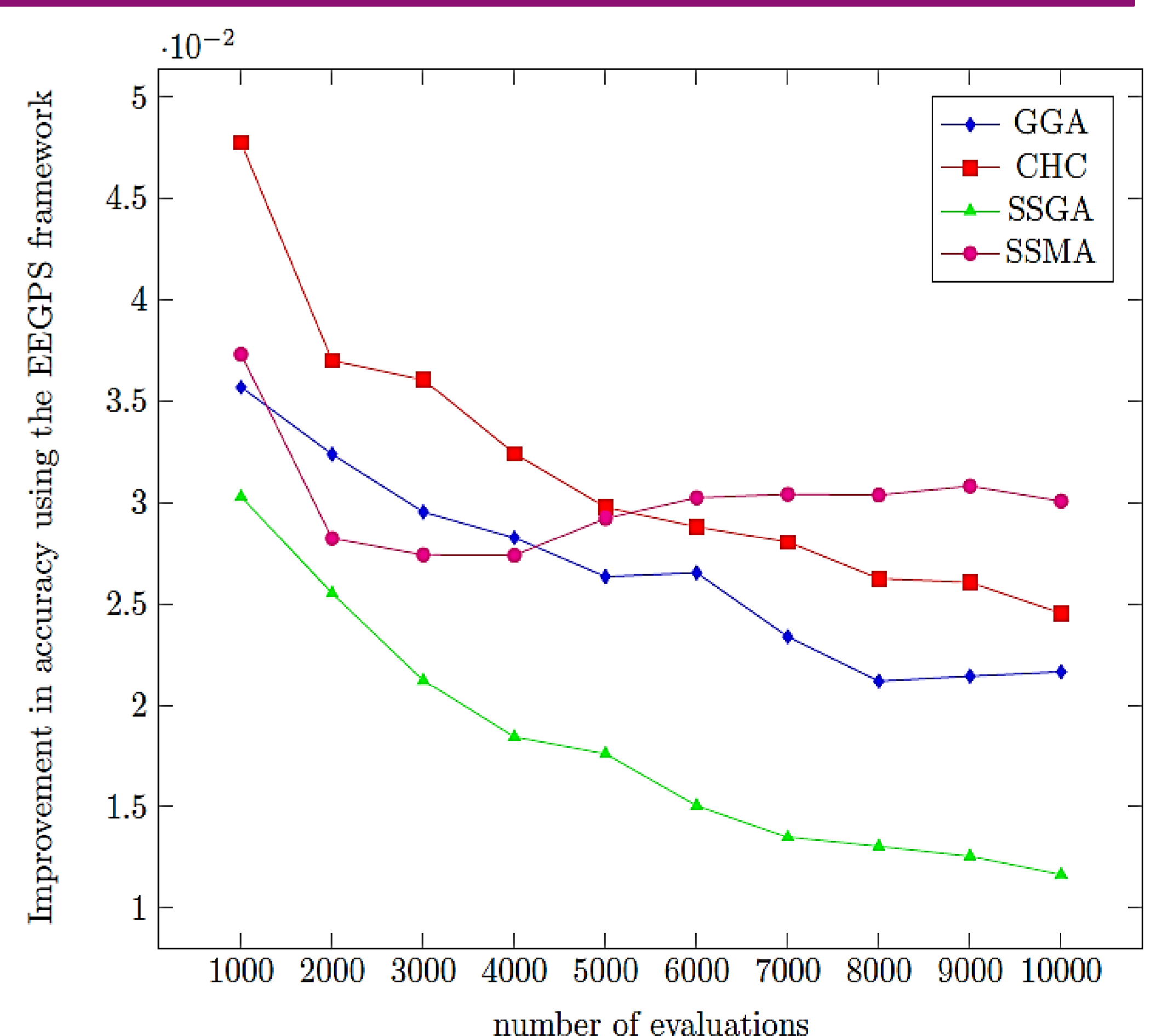
Ensemble members are weighted based on their performance in the neighborhood of the test element

Summary of results

Framework evaluated for 4 genetic PS algorithms:

- Significantly better classification performance within EEGPS
- Fewer generations are needed to obtain improvements
- Increase in computational cost is negligible
- Other genetic approaches can be easily plugged in

EEGPS: intelligent use of the work performed by a genetic PS method to boost its performance at a minimal cost



Further reading: Verbiest, N., Vluymans, S., Cornelis, C., García-Pedrajas, N., Saeys, Y.: Improving Nearest Neighbor Classification using Ensembles of Evolutionary Generated Prototype Subsets. Applied Soft Computing 44, 75-88 (2016)