ICML
International Conference
On Machine Learning

UNIVERSITY OF CAMBRIDGE

ProtoGate: Prototype-based Neural Networks with Global-to-local Feature Selection for Tabular Biomedical Data

Xiangjian Jiang, Andrei Margeloiu, Nikola Simidjievski, Mateja Jamnik

Department of Computer Science and Technology
University of Cambridge

✉: xj265@cam.ac.uk

ICML
International Conference
On Machine Learning

UNIVERSITY OF
CAMBRIDGE

ProtoGate: Prototype-based Neural Networks with Global-to-local Feature Selection for Tabular Biomedical Data

Xiangjian Jiang, Andrei Margeloiu, Nikola Simidjievski, Mateja Jamnik

Department of Computer Science and Technology
University of Cambridge

: xj265@cam.ac.uk

Contents

Part 01

Background

What task are we interested in?

01 Background

Sample names	**Gene 1**	**Gene 2**	**Gene 3**	**Gene 4**	**...**	**Gene *D***
Sample 1	1	0	1	0	...	1
Sample 2	1	1	0	0	...	0
Sample 3	0	1	0	1	...	1
Sample 4	1	1	1	1	...	1
...	...	...	...	...	...	...
Sample *N*	0	1	1	1	...	1

Table 1: An example biomedical dataset of genetic mutations.

- **Task**
  - Classification on tabular biomedical data
- **Challenges**
  - High-dimensional and low-sample-size **($D \gg N$)** $\rightarrow$ curse of dimensionality
  - Heterogeneity across samples $\rightarrow$ locally important features
- **Solution**
  - Local feature selection: select informative features on an instance-wise basis

02 Background

Sample names	**Gene 1**	**Gene 2**	**Gene 3**	**Gene 4**	**…**	**Gene *D***
Sample 1	1		1		…	1
Sample 2		1		0	…	0
Sample 3	0	1	0	1	…	
Sample 4			1		…	1
...	…	…	…	…	…	…
Sample *N*	0			1	…	

Table 2: An example of local feature selection on biomedical dataset. Colored cells denote dropped features.

- **Task**
  - Classification on tabular biomedical data
- **Challenges**
  - High-dimensional and low-sample-size ($\boldsymbol{D \gg N}$) → feature selection
  - Heterogeneity across samples → locally important features
- **Solution**
  - Local feature selection: select informative features on an instance-wise basis

Part 02

Motivation

What are the rationales behind this study?

03 Motivation

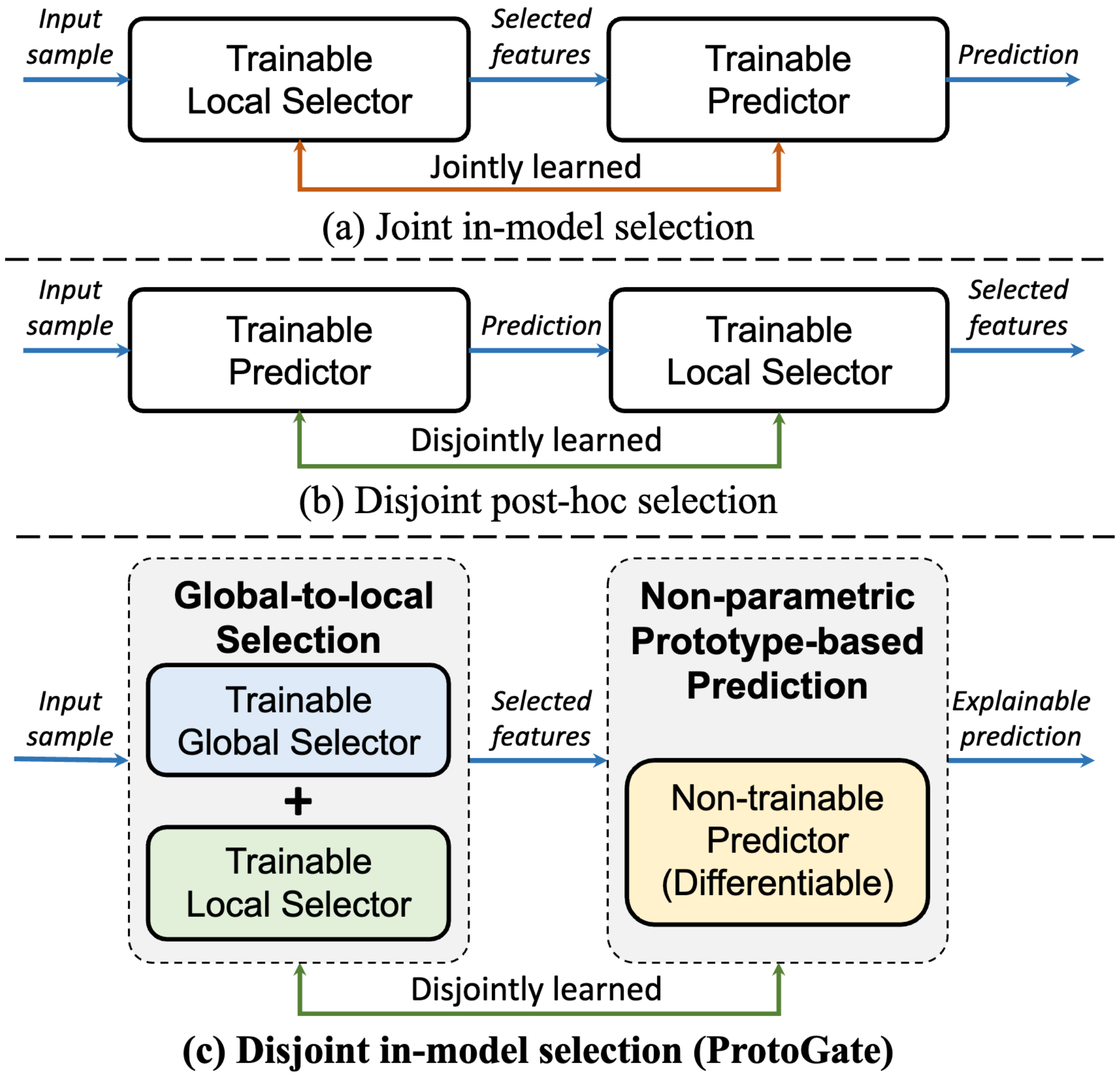


Figure 1: Overview of different paradigms.

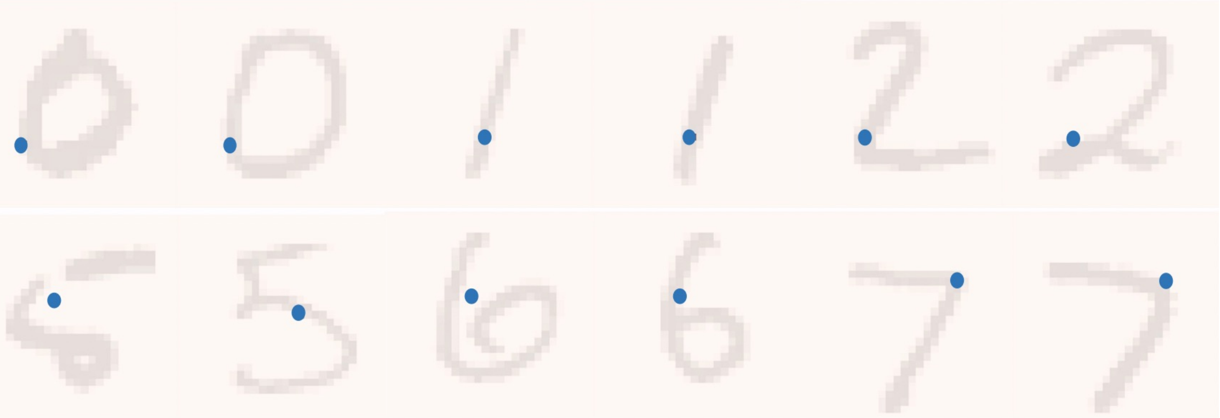


Figure 2: Illustration of co-adaptation problem.

- **Joint in-model selection**
  - ❌ susceptible to co-adaptation problem
  - ❌ insufficient explainability for predictions
- **Disjoint post-hoc selection**
  - ❌ cannot provide in-model feature importance
  - ❌ insufficient explainability for predictions
- **Disjoint in-model selection**
  - ✅ in-model feature importance
  - ✅ co-adaptation avoidance
  - ✅ human-understandable predictions

Part 03

Methodology

What is the proposed method?

Methodology | Model Design

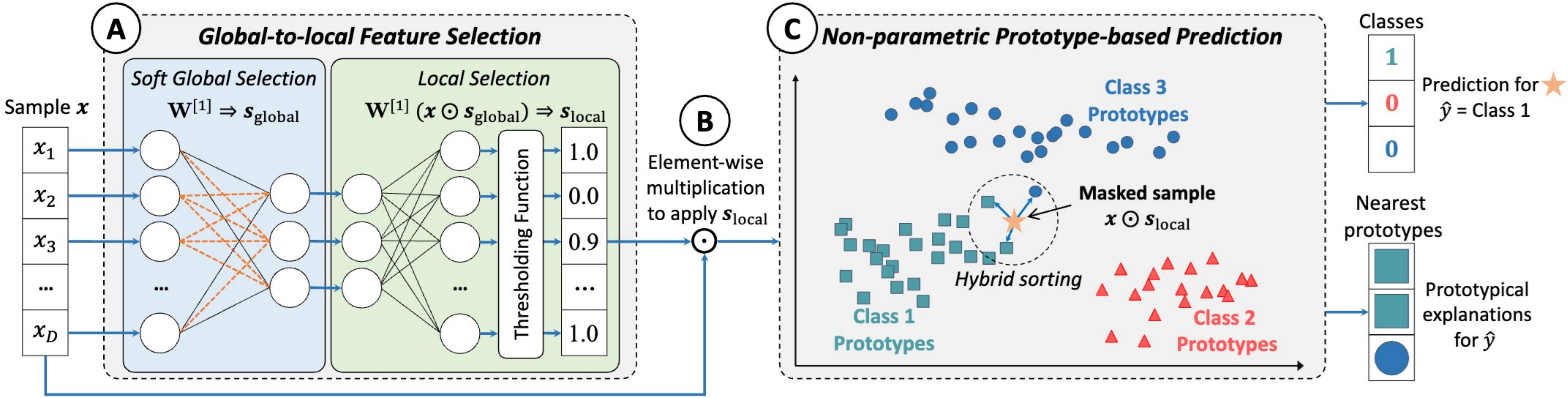


Figure 3: The model architecture of ProtoGate.

➢ Model Architecture

- Global-to-local Feature Selection *(Figure 5A)*
  - Soft global selection highlights globally important features
- Non-parametric Prototype-based Prediction *(Figure 5C)*
  - Differentiable prototype-based predictor encodes clustering assumption into selection
  - Non-parametric predictor mitigates co-adaptation problem
  - Prototypical explanations provides explainable predictions

05 Methodology | Example inference on the colon dataset

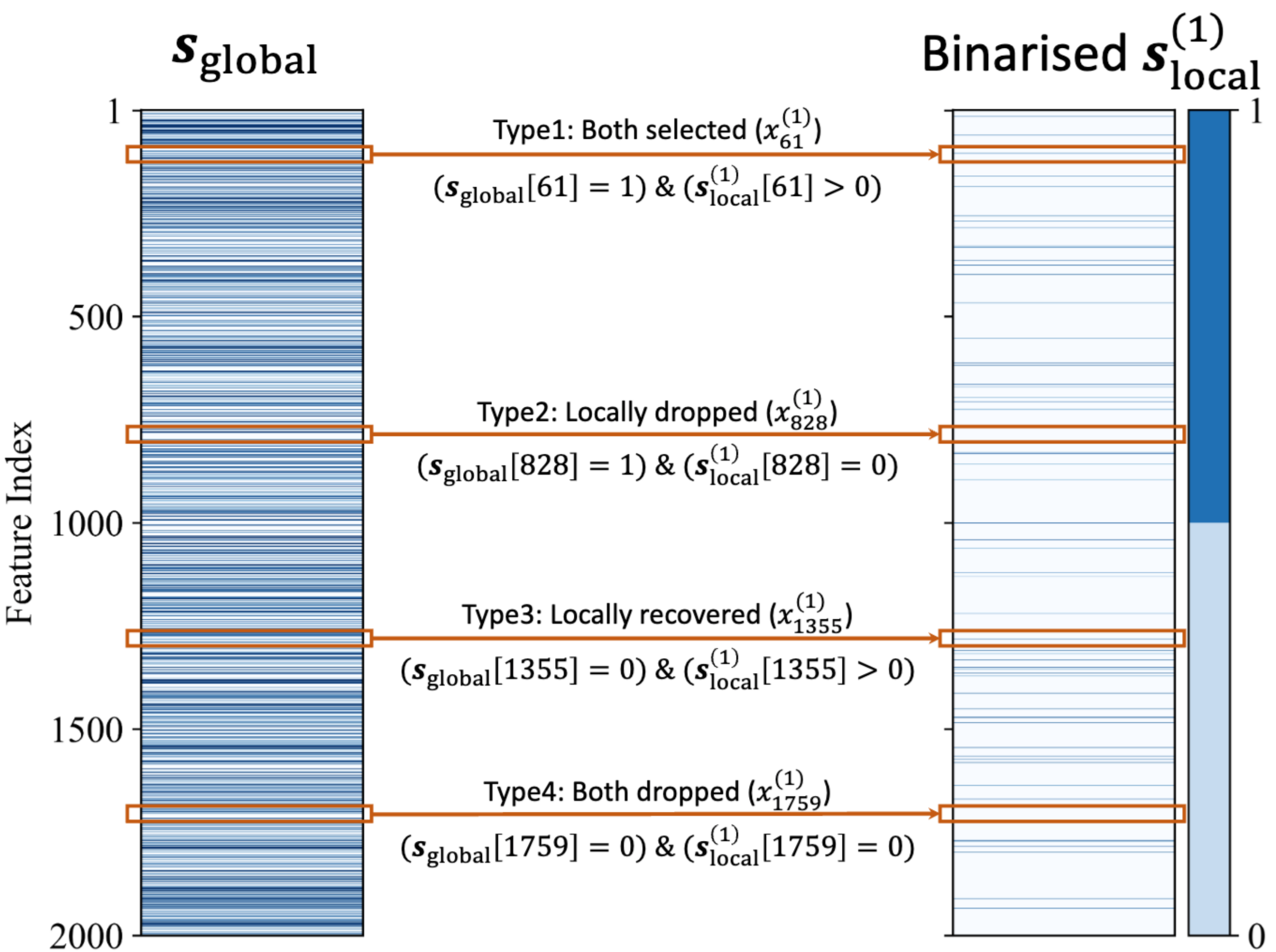
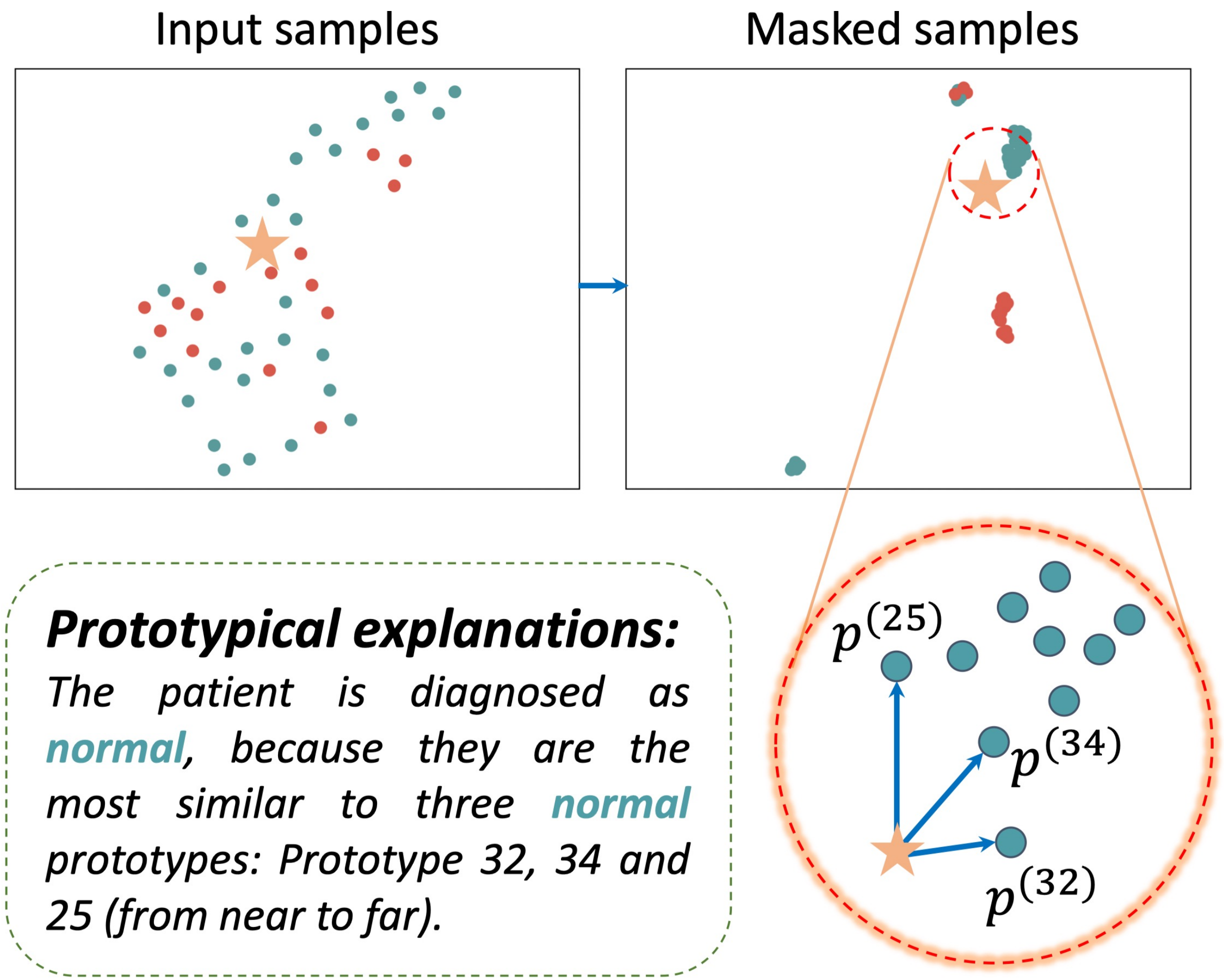


Figure 4: Illustration of the global-to-local feature selection.



Prototypical explanations:

*The patient is diagnosed as **normal**, because they are the most similar to three **normal** prototypes: Prototype 32, 34 and 25 (from near to far).*

Figure 5: Illustration of the prototype-based prediction.

Part 04

Experiments

What has been done to evaluate ProtoGate?

06 Experiments

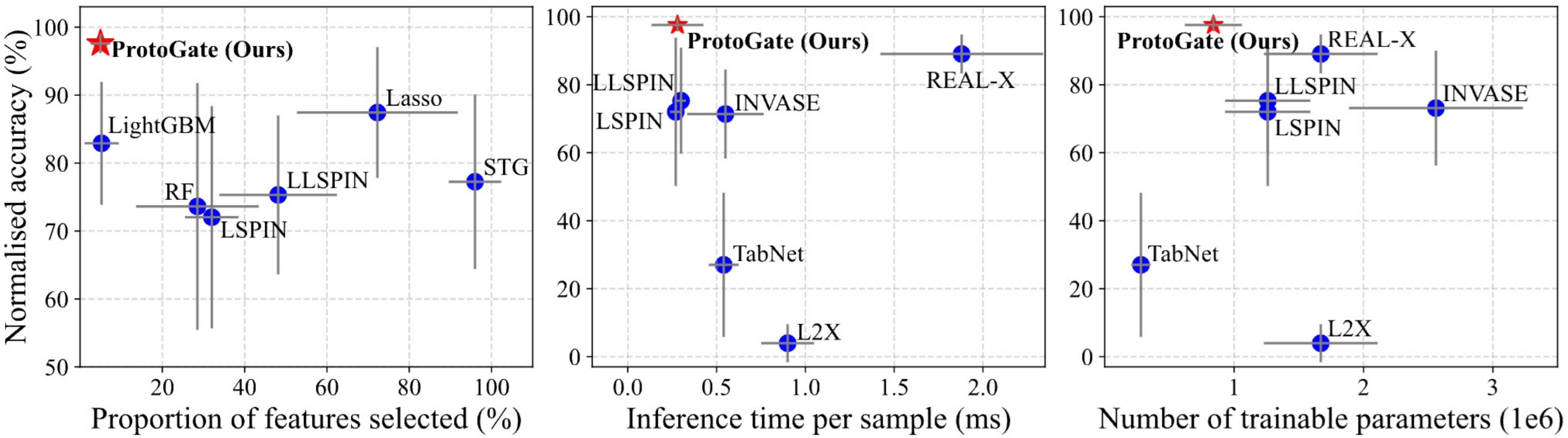


Figure 6: Predictive performance evaluation on seven real-world high-dimensional and low-sample-size datasets.

➢ Results

- ProtoGate achieves **higher accuracy** and **sparser selection** with **higher computation efficiency**
- ProtoGate selects features with **a better trade-off for fidelity**
- ProtoGate provides **easy-to-interpret prototypical predictions**, which resembles human behaviour

07 Experiments

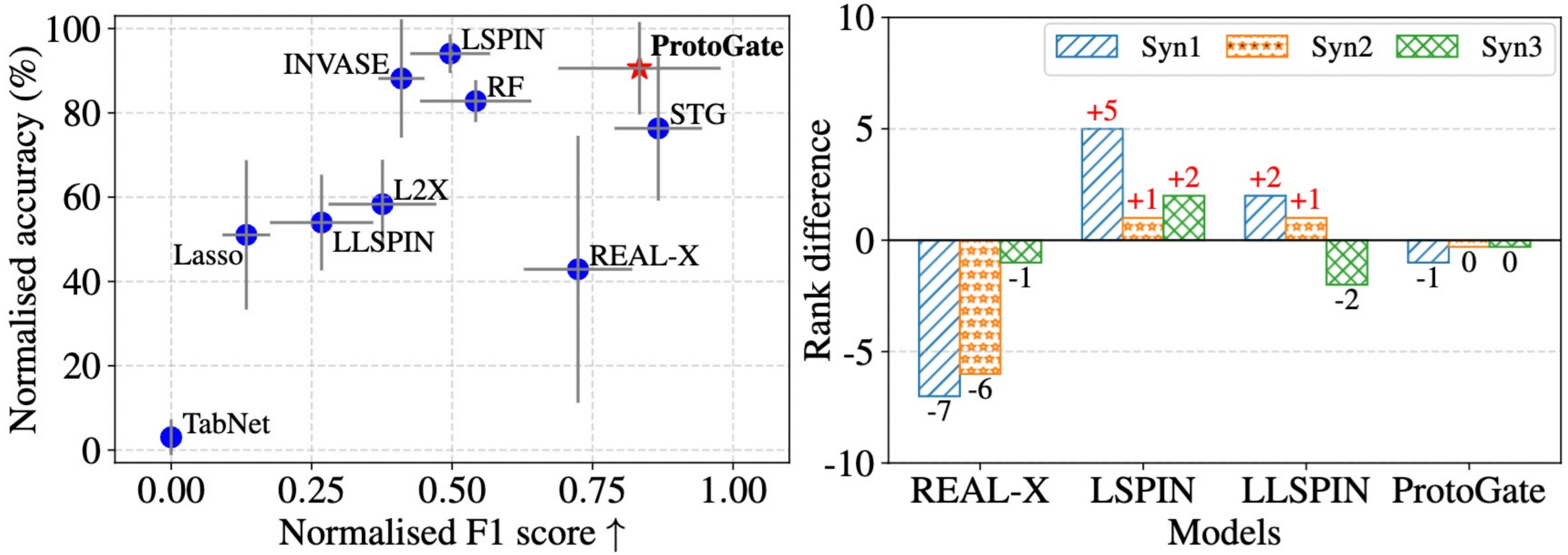


Figure 7: Fidelity evaluation of selected features on three synthetic datasets. **"Rank difference"** refers to the difference between the ranks of $F1_{select}$ (feature selection correctness) and ACC_{pred} (classification accuracy).

➢ Results

- ProtoGate achieves **higher accuracy** and **sparser selection** with **higher computation efficiency**
- ProtoGate selects features with **a better trade-off for fidelity**
- ProtoGate provides **easy-to-interpret prototypical predictions**, which resembles human behaviour

Thanks

For more details, please refer to our paper and code!
Or reach out via xj265@cam.ac.uk 🥳

📜 Paper

🖥️ Code