# Different Horses for Different Courses: Comparing Bias Mitigation Algorithms in ML

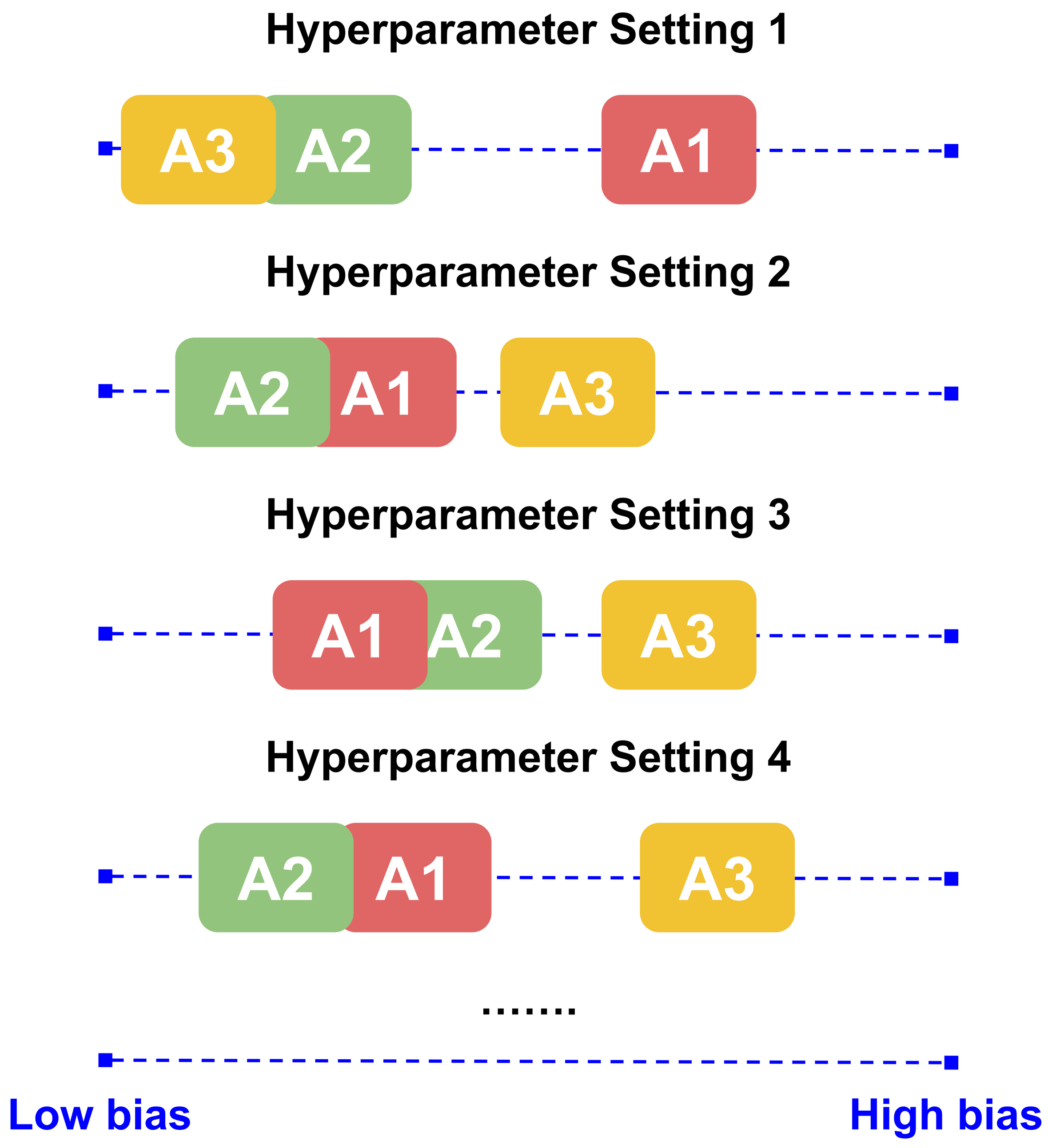
Prakhar Ganesh *[1,2] Usman Gohar *[3] Lu Cheng [4] Golnoosh Farnadi [1,2]

[1]McGill University [2]Mila [3]Iowa State University [4]University of Illinois Chicago

*equal contribution

SCAN ME

## Is a uniform evaluation setup truly 'fair'?




What is the best method if the developer is constrained to hyperparameter setting 3?

A1

What is the best method if the developer does not know the hyperparameter setting in advance, and thus wants the most robust technique?

A2

What is the best method if the developer can use any hyperparameter?

A3

What is the best method if …

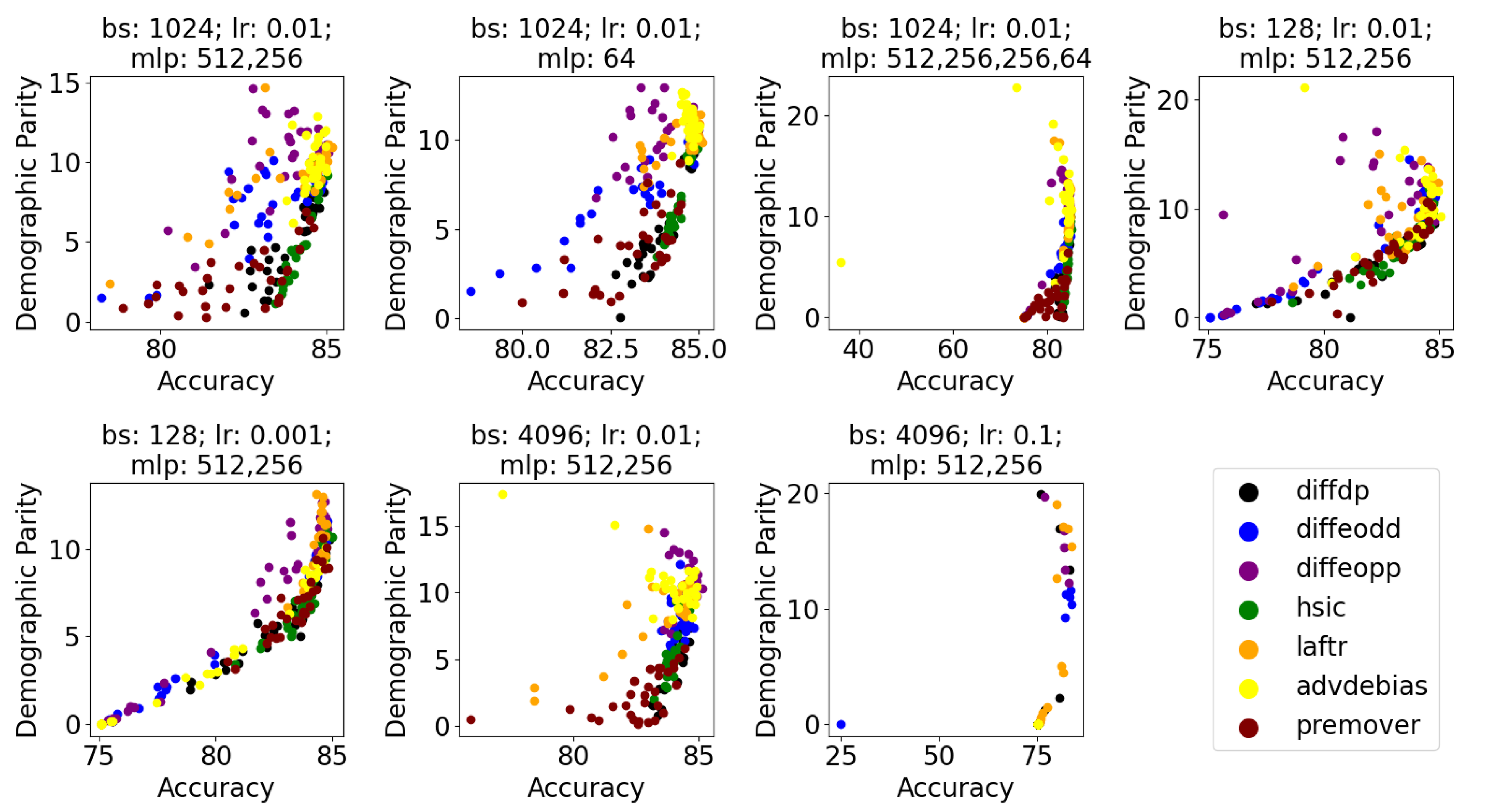
?

No method is the "best" method in every context!

No method is the "best" method in every context!

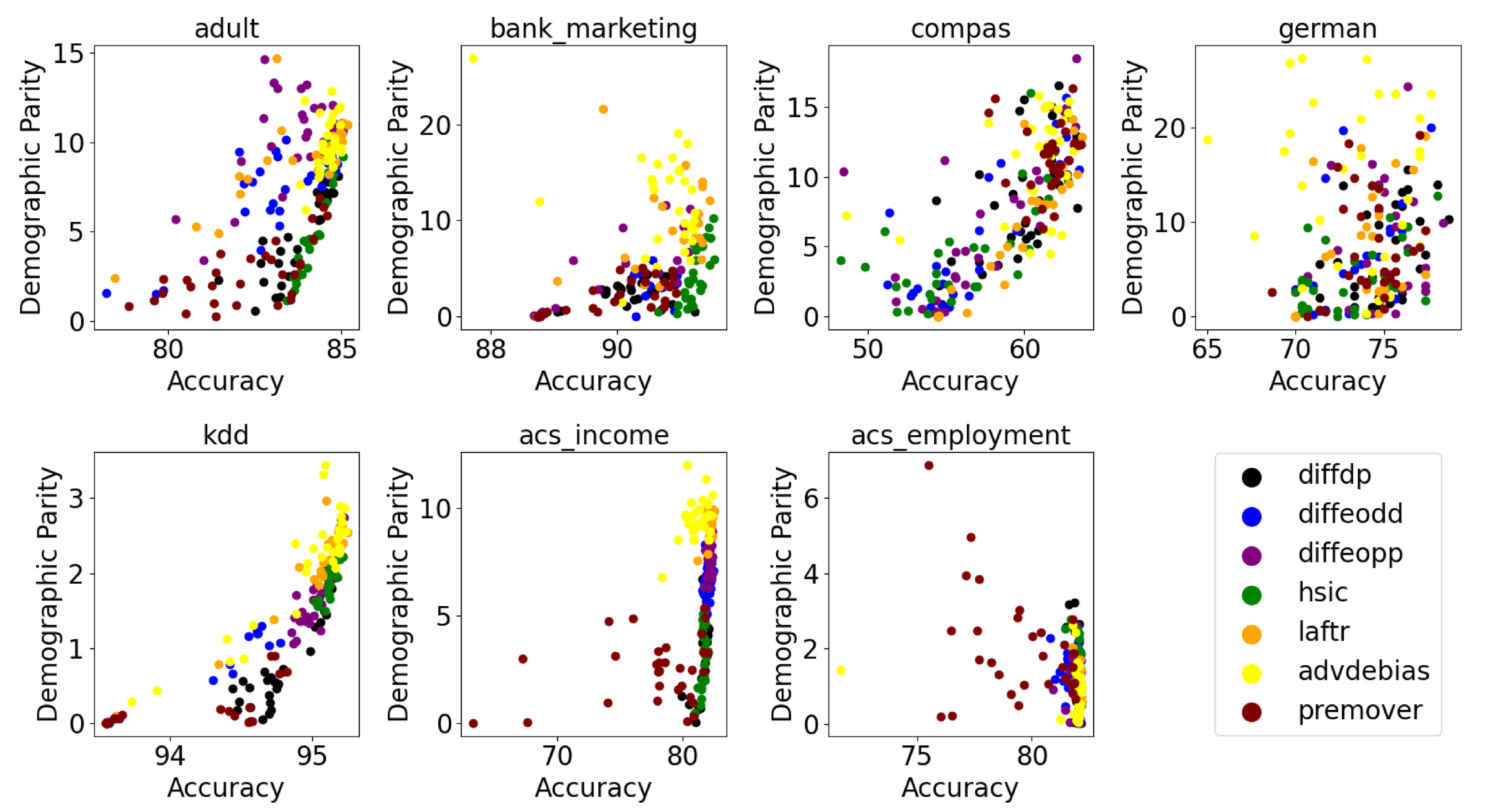
## Across Hyperparameters

A comparative analysis limited to just one hyperparameter setting fails to capture the competitive performance of all algorithms.



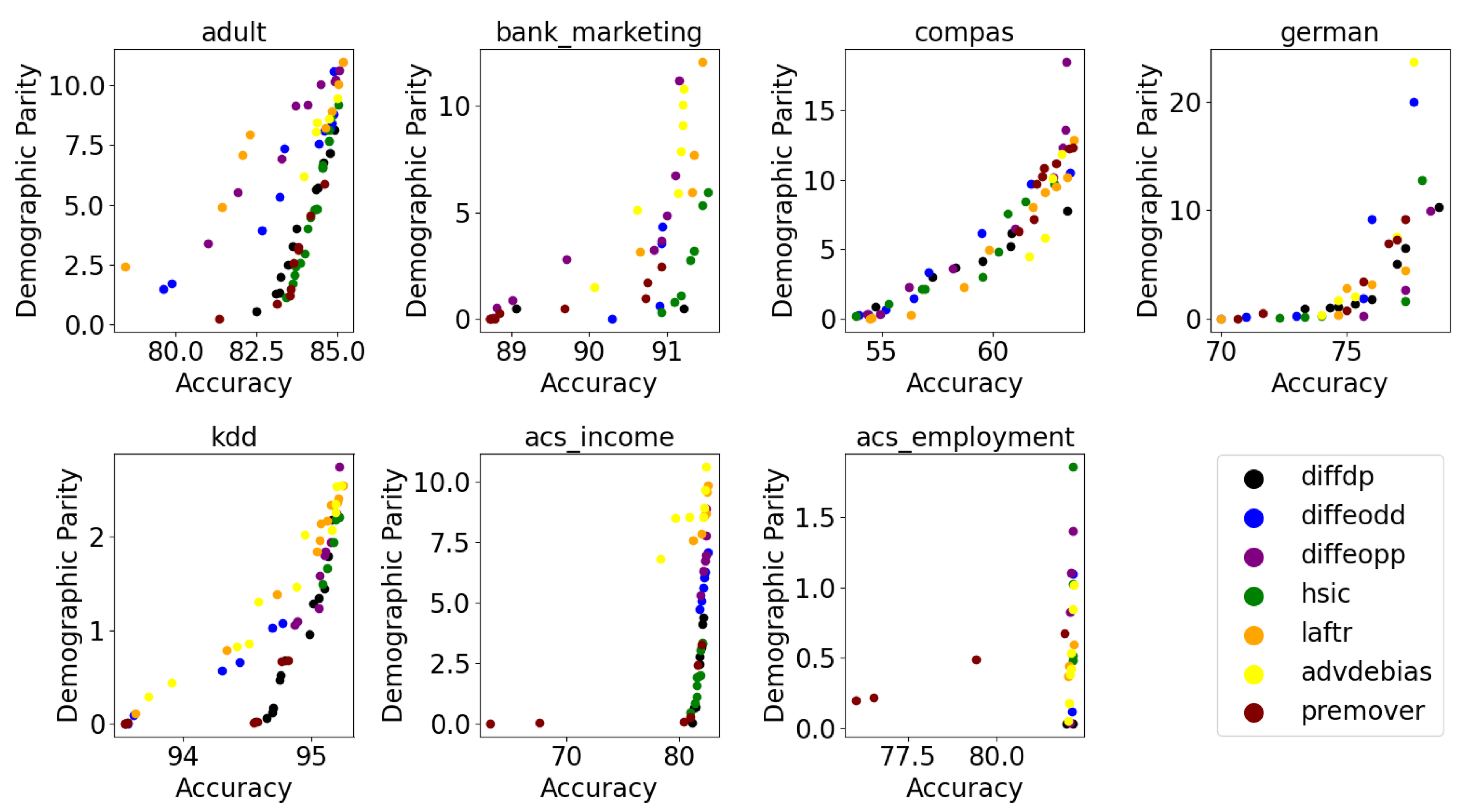

## Across Datasets

Combining results from multiple datasets can overshadow unique and rare trends, effectively hiding the shortcomings of certain algorithms.




## At the Pareto Front

Given the opportunity to perform hyperparameter optimization, most mitigation algorithms can provide competitive models!




## Benchmarking Bias Mitigation

- Current one-dimensional approaches to benchmarking are insufficient. Given the variability in fairness scores, the trends will be highly sensitive to the context. Thus, clarify the constraints of model deployment before benchmarking mitigation techniques.
- When hyperparameter optimization is possible, selecting the best technique can involve other meta-objectives beyond fairness and utility, like runtime, multiplicity, scalability constraints, etc.

## Future Work

- Our experiments were limited to in-processing techniques in bias mitigation. In the future, we will expand to pre- and post-processing.
- Our results were limited to hyperparameters during training. Further work on a large-scale study of various choices like data processing, evaluation metrics, etc., in an algorithm's lifetime is needed.

prakhar.ganesh@mila.quebec