M²FGB: A Min-Max Gradient Boosting Framework for Subgroup Fairness

Jansen S. B. Pereira

Giovani Valdrighi

Marcos Medeiros Raimundo

Universidade Estadual de Campinas,

26 June 2025
ACM FAccT

ACM FACCT
ΑΘΗΝΑ 2025

UNICAMP

H.IAAC HUB DE INTELIGÊNCIA ARTIFICIAL E ARQUITETURAS COGNITIVAS

Summary

- We consider the problem of **min-max fairness:** improving the condition of the most disadvantaged subgroup.
- Novel framework built upon **gradient-boosting** that supports **multiple fairness criteria** and **multiple subgroups**.
- Our solution **converges** and presents **positive results in experiments** with benchmark datasets.

Two views of fairness in supervised learning

-

$$\min_{f} \quad \mathcal{L}(f)$$ minimize a loss of model f

Two views of fairness in supervised learning

-

$$\min_{f} \quad \mathcal{L}(f)$$

$$s.t.\ |\mathcal{L}_z(f) - \mathcal{L}_{z'}(f)| \leq c$$

Disparity based fairness

Two views of fairness in supervised learning

-

$$\min_{f} \quad \mathcal{L}(f)$$

$$s.t.\ |\mathcal{L}_z(f) - \mathcal{L}_{z'}(f)| \leq c$$

Disparity based fairness

$$\min_{f} \quad \max_{\underline{z \in \mathcal{Z}}} \underline{\mathcal{L}_z}(f)$$

Min-max fairness

subgroups

group-wise loss

Prior work

MMPF (Martinez et al. 2020)

Min-max fairness as a multi-objective optimization problem

MinMaxFair (Diana et al. 2021)

Min-max fairness as a game theory problem

Prior work

MMPF (Martinez et al. 2020)

Min-max fairness as a multi-objective optimization problem

✅ Min-max fairness
❌ Costly iterative algorithm

MinMaxFair (Diana et al. 2021)

Min-max fairness as a game theory problem

✅ Min-max fairness
❌ Costly iterative algorithm

Prior work

MMPF (Martinez et al. 2020)

Min-max fairness as a multi-objective optimization problem

✅ Min-max fairness
❌ Costly iterative algorithm

MinMaxFair (Diana et al. 2021)

Min-max fairness as a game theory problem

✅ Min-max fairness
❌ Costly iterative algorithm

FairGBM (Cruz et al. 2023)

Gradient-boosting with fairness constraints

Prior work

MMPF (Martinez et al. 2020)

Min-max fairness as a multi-objective optimization problem

✅ Min-max fairness
❌ Costly iterative algorithm

MinMaxFair (Diana et al. 2021)

Min-max fairness as a game theory problem

✅ Min-max fairness
❌ Costly iterative algorithm

FairGBM (Cruz et al. 2023)

Gradient-boosting with fairness constraints

❌ Disparity based fairness
✅ Gradient-boosting efficiency

Updated problem

- Weighted objective with a fairness strength parameter λ

$$\min_f \left[(1-\lambda)\mathcal{L} + \lambda \max_z \{\mathcal{L}_z\} \right]$$

overall loss

group-wise loss

Updated problem

- Weighted objective with a fairness strength parameter λ

$$\min_f \left[(1-\lambda)\underbrace{\mathcal{L}}_{\text{overall loss}} + \lambda \max_z \{\underbrace{\mathcal{L}_z}_{\text{group-wise loss}}\}\right]$$

- Overall loss and group-wise loss can be different

Updated problem

- Weighted objective with a fairness strength parameter λ

$$\min_f \left[(1-\lambda)\underbrace{\mathcal{L}}_{\text{overall loss}} + \lambda \max_z \{ \underbrace{\mathcal{L}_z}_{\text{group-wise loss}} \} \right]$$

- Overall loss and group-wise loss can be different
- The Lagrangian introduces dual variables μ_z

$$L = (1-\lambda)\mathcal{L} + \lambda \sum_z \underbrace{\mu_z}_{\text{“weight” on each objective}} \mathcal{L}_z + \epsilon \left(\lambda - \sum_z \mu_z \right)$$

Algorithm

- We leverage gradient-boosting to optimize this problem

$$\underbrace{f^{t+1}}_{\text{ensemble}} = f^t + \gamma h \qquad \underbrace{h \approx -\nabla L(f^t)}_{\text{weak approximation of gradient}}$$

- **Primal-dual** optimization between boosting rounds

Algorithm – in simple steps

1. Weights are initiliazed with equal values

Algorithm – in simple steps

1. Weights are initiliazed with equal values
2. For M iterations:
    1. f is updated using gradient descent

$$h \approx -\nabla\Big((1-\lambda)\underbrace{\mathcal{L}}_{\text{overall loss}} + \lambda\Sigma_z \underbrace{\mu_z \mathcal{L}_z}_{\text{weighted group-wise loss}}\Big)$$

Algorithm – in simple steps

1. Weights are initiliazed with equal values
2. For M iterations:
    1. f is updated using gradient descent

    $$h \approx -\nabla\Big((1-\lambda)\underbrace{\mathcal{L}}_{\text{overall loss}} + \lambda\Sigma_z \underbrace{\mu_z \mathcal{L}_z}_{\text{weighted group-wise loss}}\Big)$$

    2. μ_z is updated using projected gradient ascent

    $$\mu_z \leftarrow \Pi(\mu_z + \alpha\mathcal{L}_z)$$

    high loss → high weight

Algorithm – in simple steps

1. Weights are initiliazed with equal values
2. For M iterations:
   1. f is updated using gradient descent

$$h \approx -\nabla\left((1-\lambda)\underline{\mathcal{L}} + \lambda\Sigma_z \underline{\mu_z \mathcal{L}_z}\right)$$

overall loss weighted group-wise loss

   2. μ_z is updated using projected gradient ascent

$$\mu_z \leftarrow \Pi(\mu_z + \alpha\mathcal{L}_z)$$

high loss → high weight

Converges if learning rate is sufficiently small and $\langle \nabla_f L(f^t, \mu^t), h^t \rangle \leq 0$

Fairness metrics

- $\mathcal{L}_z$ can be any **differentiable** function
- It evaluates the "quality" of model f for group z

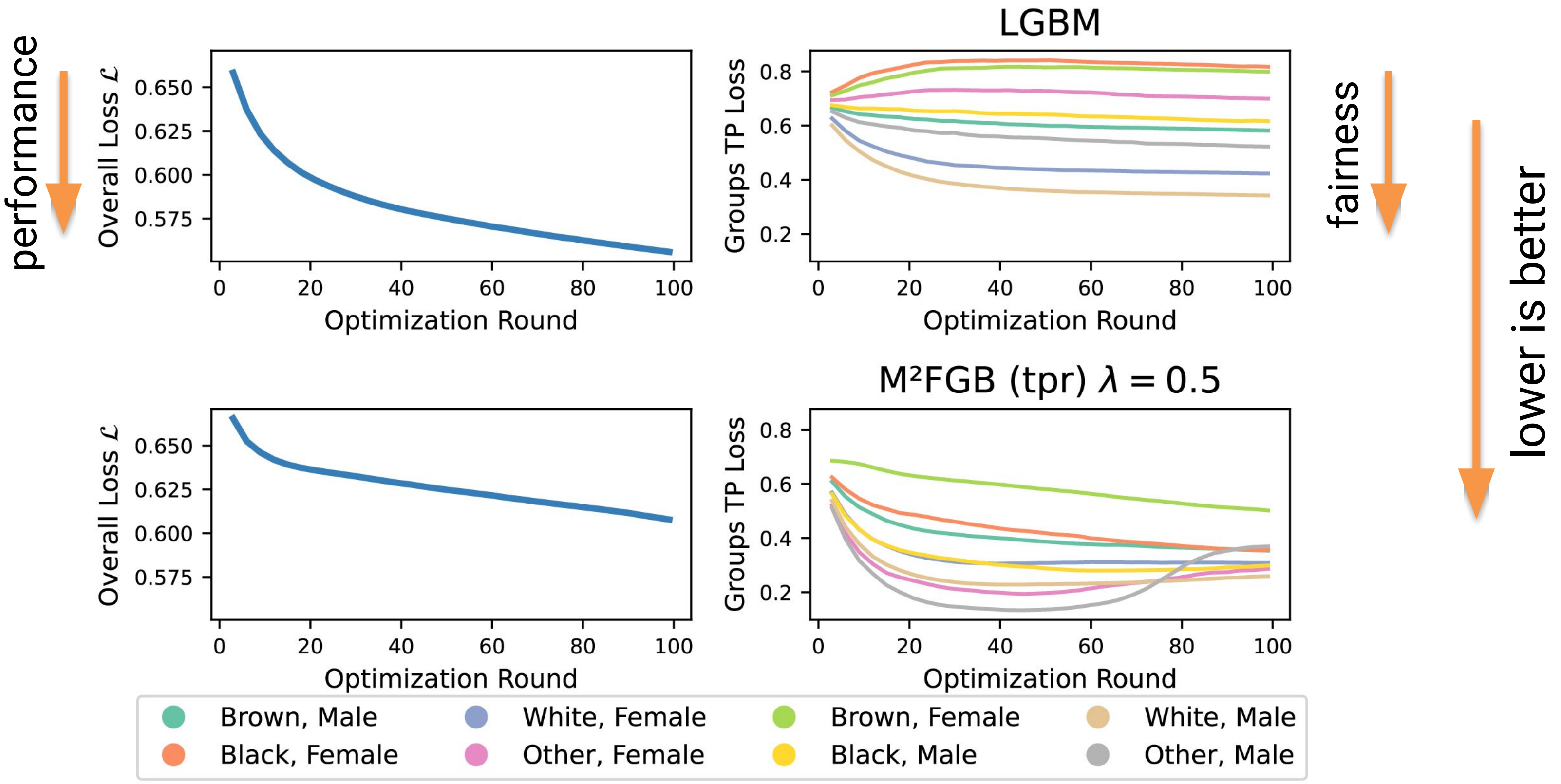
- We present $\mathcal{L}_z$ based on:
  - Equality of opportunity
  - Equality of accuracy
  - Demographic parity
- Adaptations of binary cross entropy

Experiments

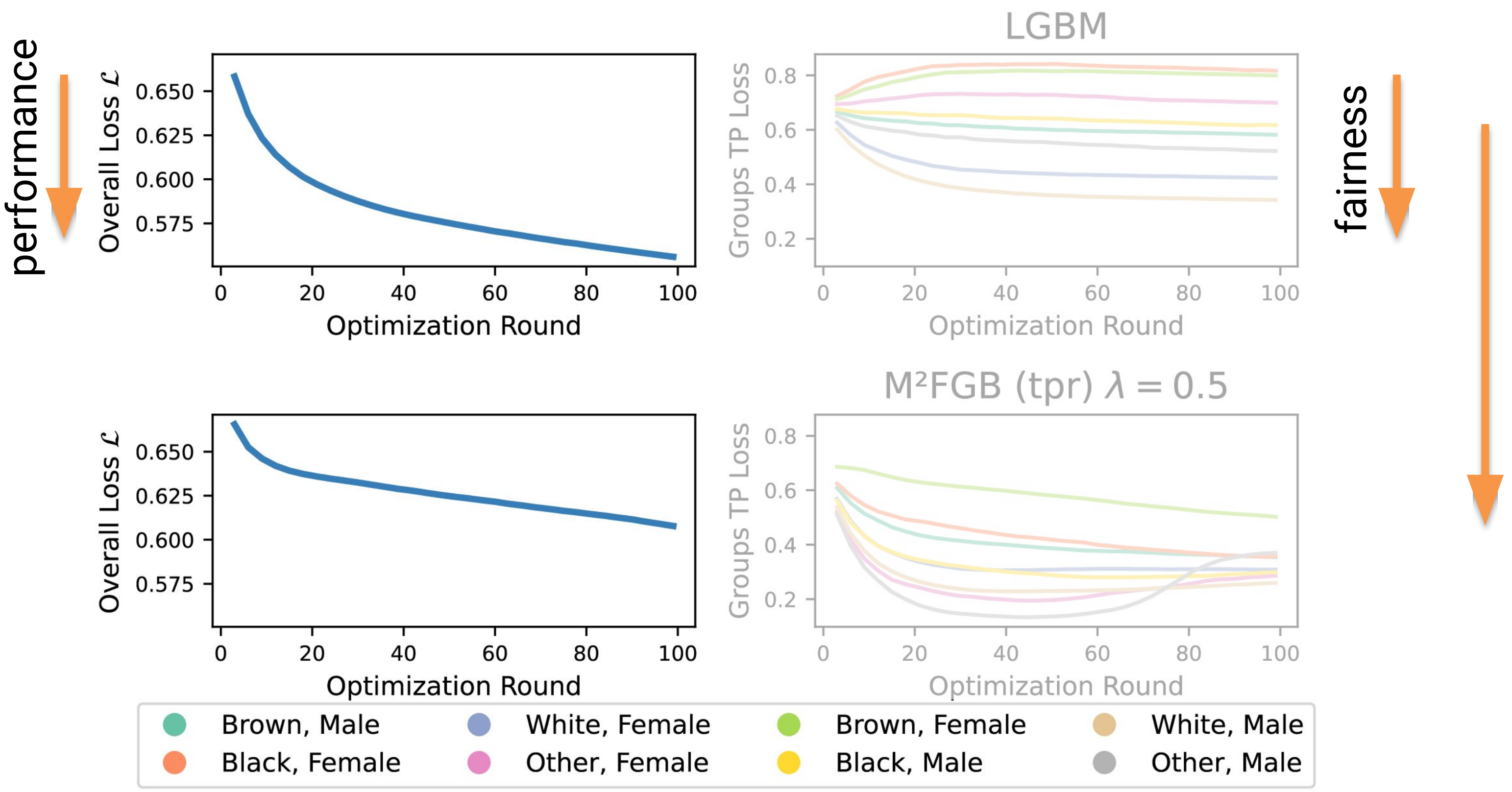
- Benchmark on 4 tabular datasets
  - From 1000 samples to 1M
  - Sensitive subgroups based in race, gender and age
  - 4 or 8 subgroups
- Comparison with
  - Gradient-boosting baseline
  - Prior min-max fair algorithms

Check the paper for all results!

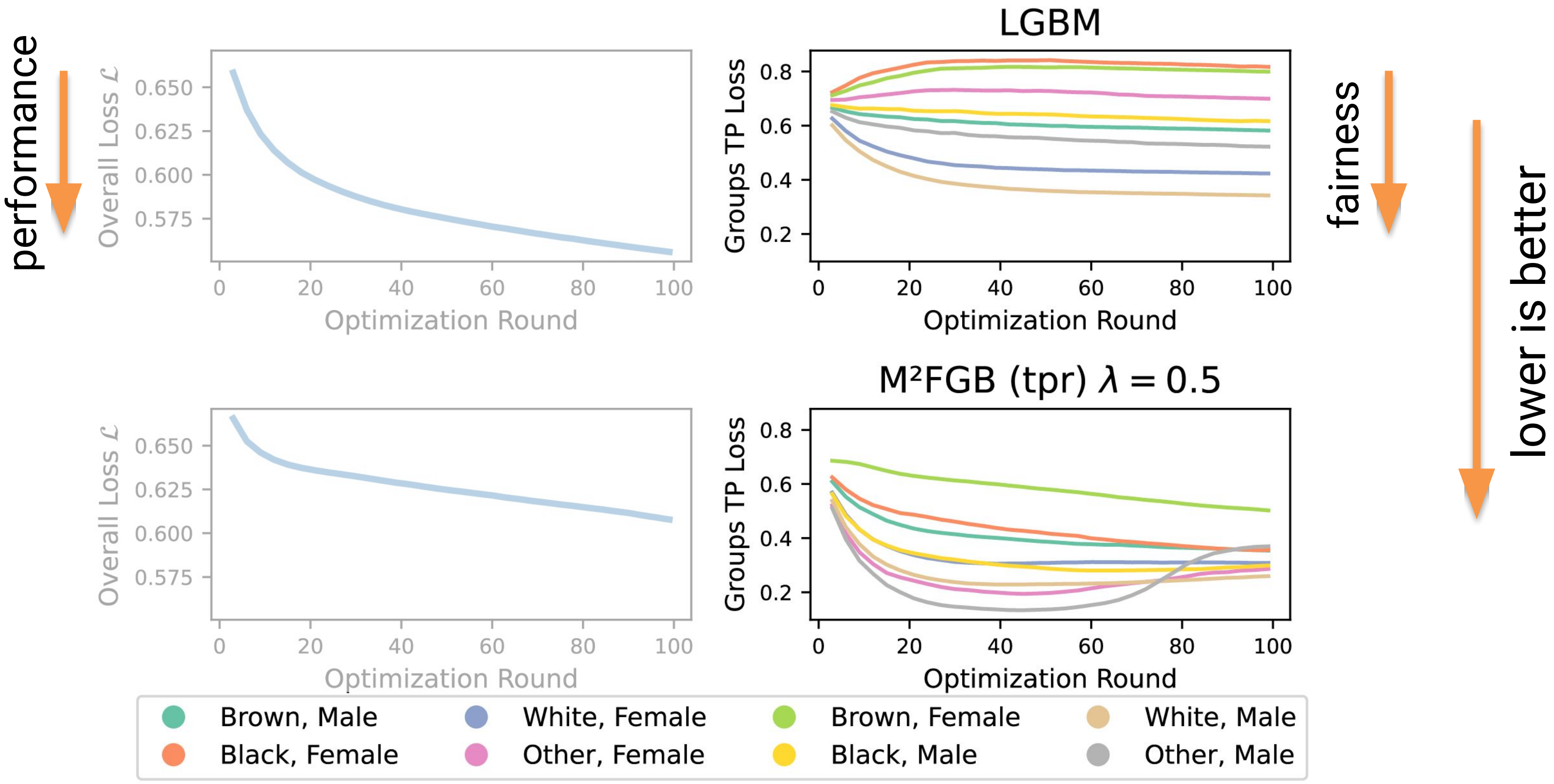
Evaluation against gradient-boosting



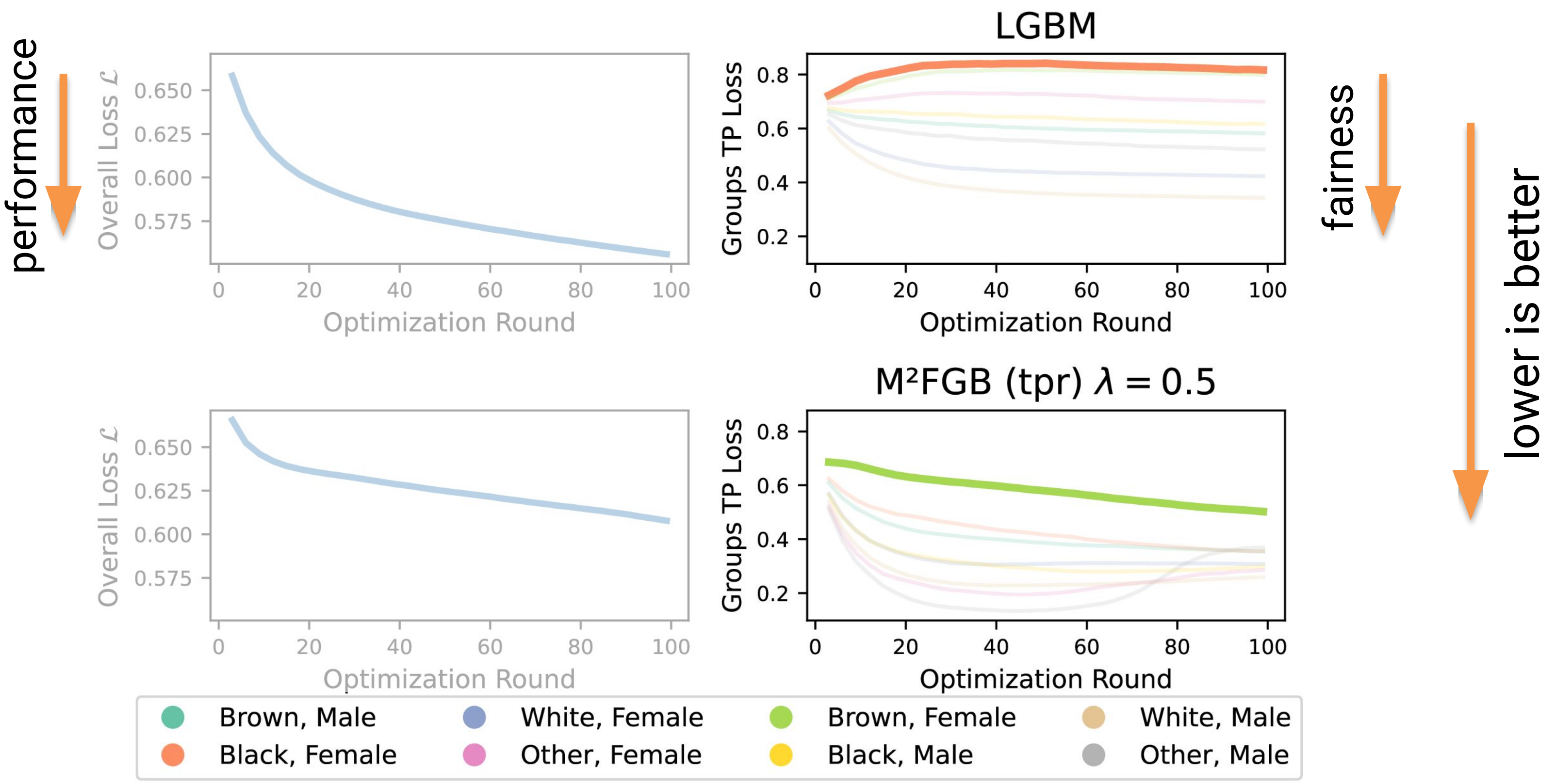
Evaluation against gradient-boosting



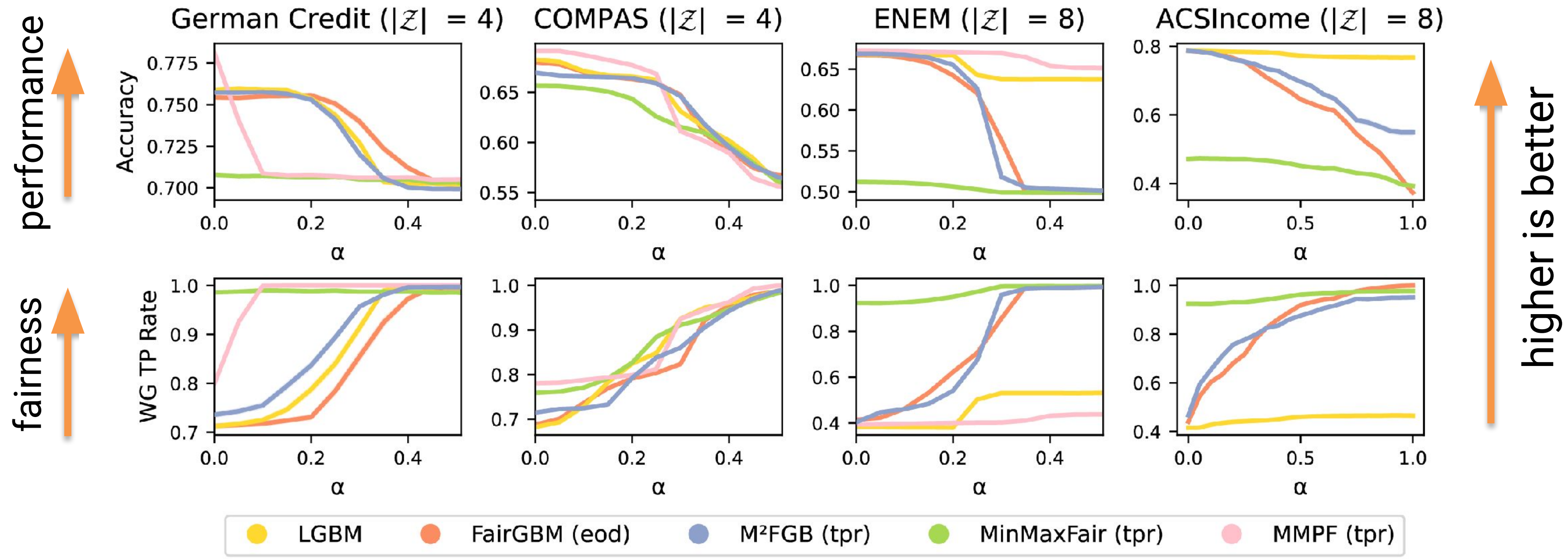
Evaluation against gradient-boosting



Evaluation against gradient-boosting



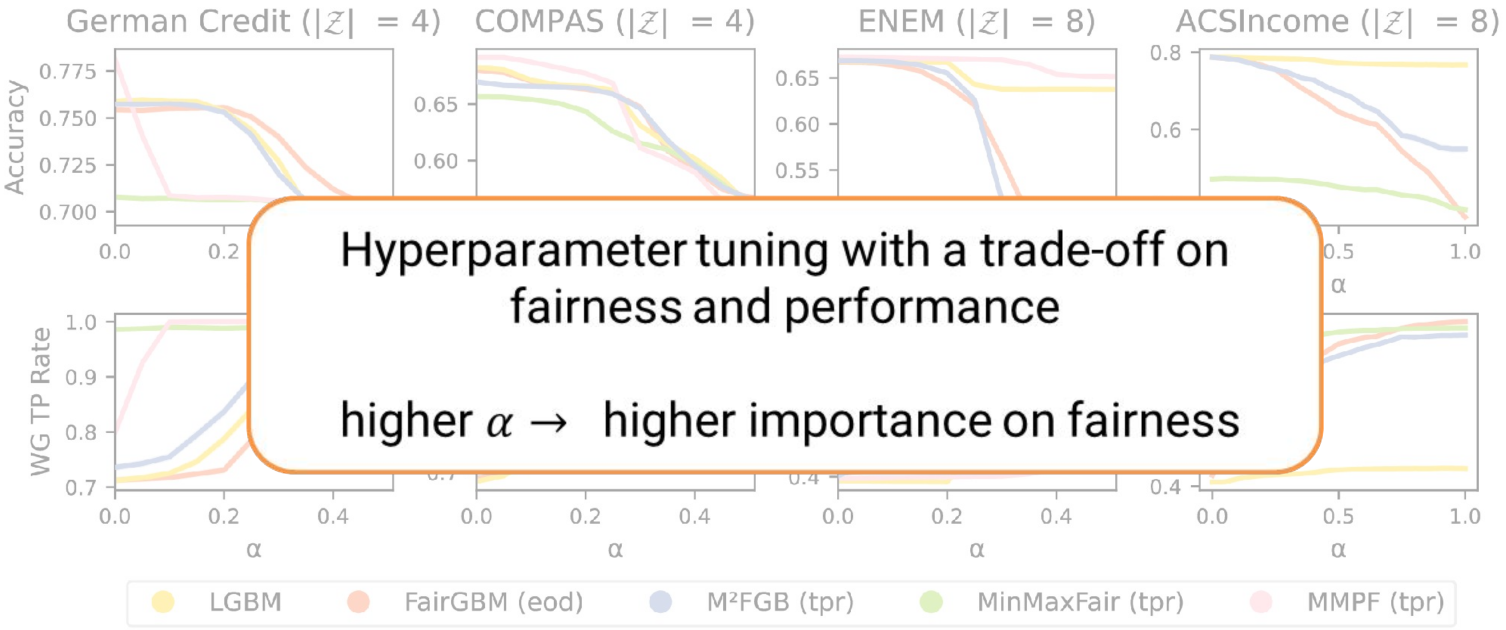
Comparison with hyperparameter tuning



Comparison with hyperparameter tuning

performance

fairness

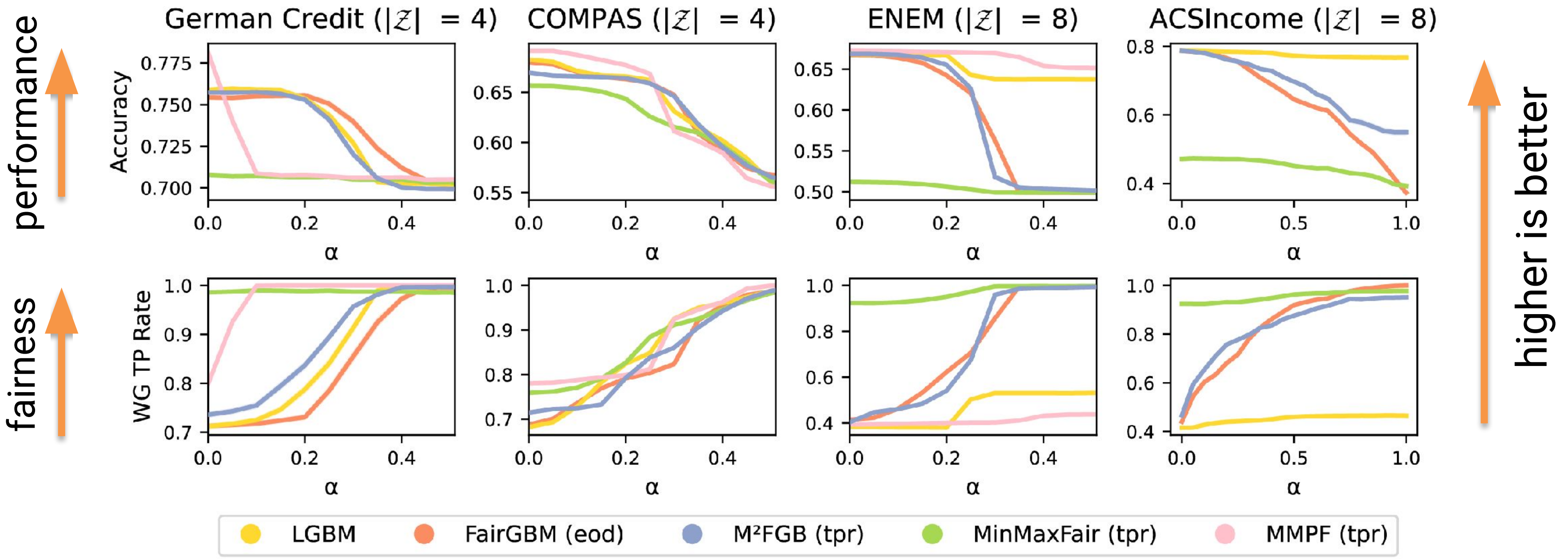


Hyperparameter tuning with a trade-off on fairness and performance

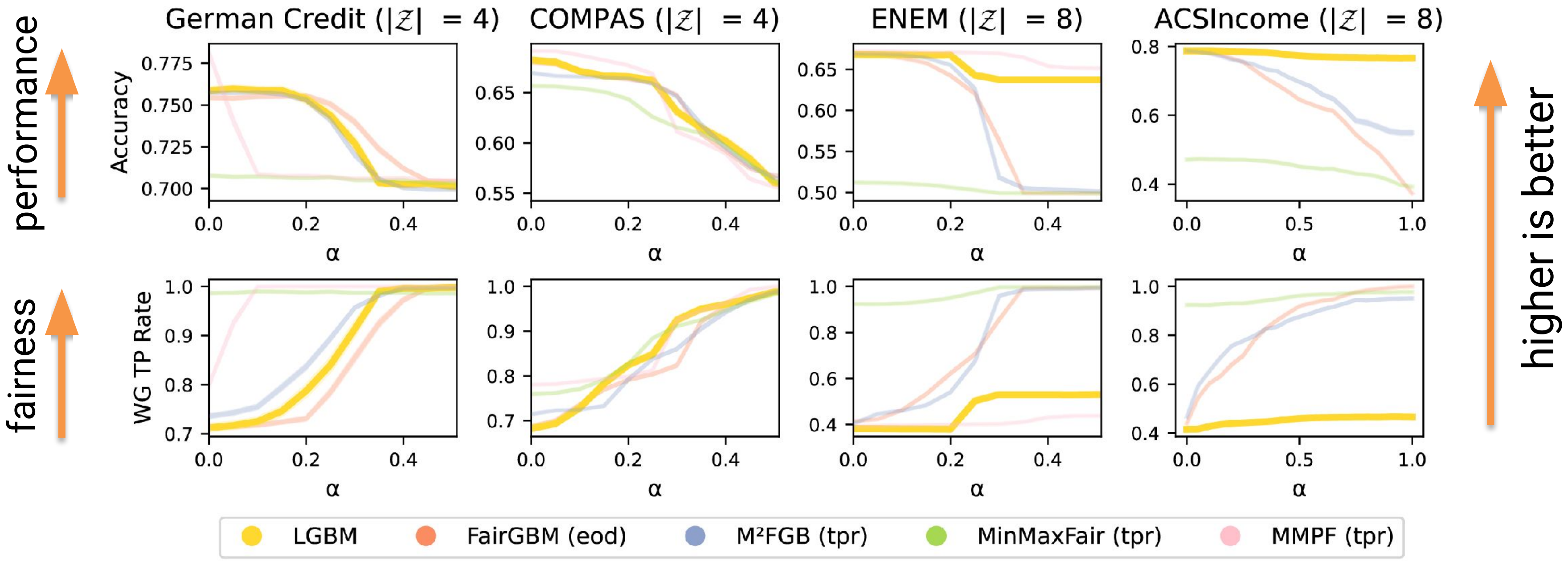
higher $\alpha \rightarrow$ higher importance on fairness

higher is better

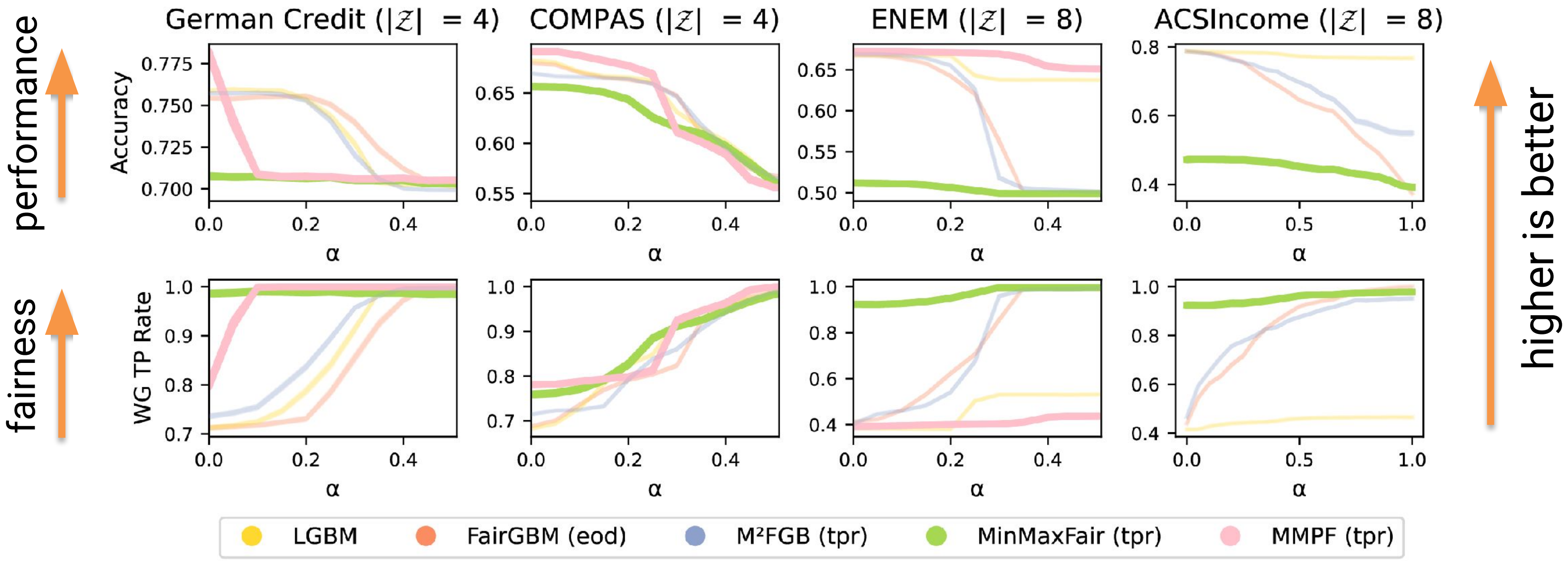
Comparison with hyperparameter tuning



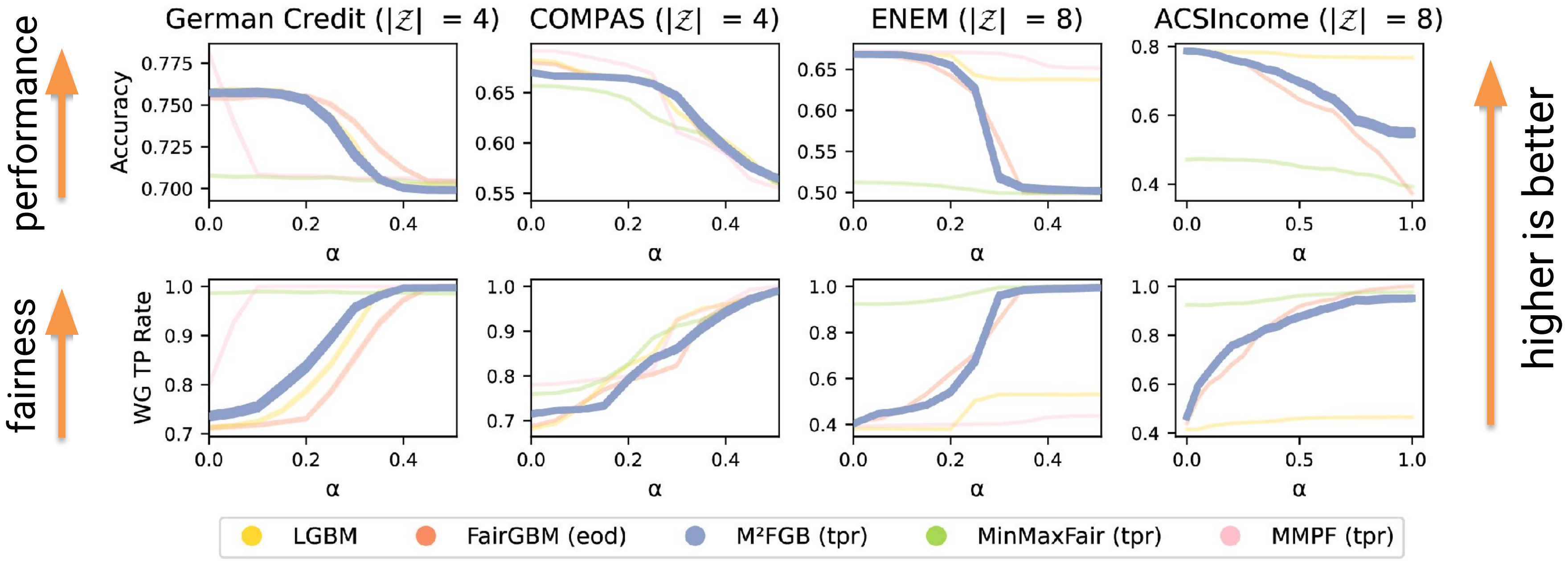
Comparison with hyperparameter tuning



Comparison with hyperparameter tuning



Comparison with hyperparameter tuning



Improved fairness without significant decrease in performance

Computing time comparison

Model	German	COMPAS	ENEM	ACSIncome
LGBM	1	1	1	1
FairGBM	x0.9	x2.2	x0.6	X1.0
MMPF	x27.8	x19.1	x11.4	---
MinMaxFair	x15.6	x42.1	x11.9	x24.0
M^2FGB	x4.1	x2.4	x1.7	x2.3

Computing time as a ratio of LGBM computing time.

Limitations

- A large number of boosting rounds might improve fairness, but it may also increase overfitting. It is important to perform hyperparameter optimization.

Limitations

- A large number of boosting rounds might improve fairness, but it may also increase overfitting. It is important to perform hyperparameter optimization.

Future works

- Study generalization of groups to end training and avoid overfitting.
- Adapt algorithm to other data formats, such as images and text.

Summary

- We present **M^2FGB**, a framework based on **gradient-boosting** that optimizes for **min-max fairness**.
- Our algorithm is capable of working with **multiple fairness loss** functions and **multiple groups**.
- M^2FGB **converges** and has shown significant improvement in **efficiency** with big datasets.

M²FGB: A Min-Max Gradient Boosting Framework for Subgroup Fairness

Paper

Code

e-mail: giovani.valdrighi@ic.unicamp.br

Thanks! Questions?

ACM FACCT
ΑΘΗΝΑ 2025

UNICAMP

H.IAAC HUB DE INTELIGÊNCIA ARTIFICIAL E ARQUITETURAS COGNITIVAS