Appendix: Detailed Results

The following tables contain the detailed performance of all algorithms separated by dataset. These results were aggregated to create the rankings displayed in the Experimental Evaluation. Results contain the mean and standard error across all 10 trials.

Table A1: Detailed results for **ADULT - SEX MALE**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.69 ± 0.02	0.44 ± 0.02	0.78 ± 0.01	0.38 ± 0.00	0.54 ± 0.05	0.07 ± 0.01	0.05 ± 0.01	0.00 ± 0.00	0.46 ± 0.02
Hardt	0.70 ± 0.00	0.41 ± 0.00	0.58 ± 0.01	0.27 ± 0.00	0.87 ± 0.00	0.06 ± 0.00	0.03 ± 0.01	0.00 ± 0.00	0.67 ± 0.02
Feldman	0.85 ± 0.00	0.55 ± 0.00	0.81 ± 0.00	0.45 ± 0.00	0.71 ± 0.00	0.23 ± 0.00	0.28 ± 0.01	0.00 ± 0.00	0.26 ± 0.03
Adv. Deb.	0.70 ± 0.00	0.45 ± 0.00	0.75 ± 0.00	0.36 ± 0.00	0.61 ± 0.01	0.42 ± 0.01	1.06 ± 0.01	0.00 ± 0.00	0.25 ± 0.01
xFAIR	0.83 ± 0.00	0.44 ± 0.00	0.83 ± 0.00	0.50 ± 0.00	0.39 ± 0.00	0.12 ± 0.00	0.18 ± 0.01	0.10 ± 0.03	0.17 ± 0.00
MLP_{F1}	0.85 ± 0.00	0.55 ± 0.00	0.81 ± 0.00	0.46 ± 0.00	0.69 ± 0.01	0.21 ± 0.00	0.24 ± 0.01	0.00 ± 0.00	0.30 ± 0.04
MLP_{SP}	0.49 ± 0.02	0.14 ± 0.05	0.50 ± 0.11	0.08 ± 0.03	0.50 ± 0.16	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.06 ± 0.01
MLP_{EDiff}	0.44 ± 0.03	0.09 ± 0.04	0.63 ± 0.10	0.05 ± 0.02	0.30 ± 0.14	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.06 ± 0.00
MLP_{F1+SP}	0.81 ± 0.01	0.55 ± 0.00	0.82 ± 0.00	0.46 ± 0.00	0.68 ± 0.00	0.20 ± 0.00	0.21 ± 0.01	0.00 ± 0.00	0.43 ± 0.03
$MLP_{F1+EDiff}$	0.84 ± 0.00	0.52 ± 0.00	0.80 ± 0.00	0.43 ± 0.00	0.68 ± 0.00	0.23 ± 0.00	0.30 ± 0.01	0.06 ± 0.05	0.12 ± 0.01
$MLP_{SP+EDiff}$	0.50 ± 0.02	0.09 ± 0.04	0.63 ± 0.10	0.05 ± 0.02	0.30 ± 0.14	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.01	0.05 ± 0.01
MLP_{All}	0.81 ± 0.00	0.52 ± 0.00	0.80 ± 0.00	0.43 ± 0.00	0.65 ± 0.01	0.15 ± 0.01	0.16 ± 0.01	0.00 ± 0.00	0.14 ± 0.00
$MLP_{Curriculum}$	0.81 ± 0.00	0.52 ± 0.00	0.80 ± 0.00	0.43 ± 0.00	0.65 ± 0.01	0.15 ± 0.01	0.16 ± 0.01	0.00 ± 0.00	0.15 ± 0.01

Table A2: Detailed results for **ADULT - RACE WHITE**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.71 ± 0.02	0.48 ± 0.02	0.80 ± 0.00	0.42 ± 0.00	0.57 ± 0.05	0.04 ± 0.00	0.06 ± 0.01	0.10 ± 0.04	0.40 ± 0.04
Hardt	0.75 ± 0.00	0.47 ± 0.00	0.66 ± 0.00	0.31 ± 0.00	0.89 ± 0.00	0.03 ± 0.01	0.05 ± 0.01	0.00 ± 0.00	0.68 ± 0.04
Feldman	0.85 ± 0.00	0.55 ± 0.00	0.79 ± 0.00	0.43 ± 0.00	0.74 ± 0.01	0.13 ± 0.01	0.16 ± 0.02	0.10 ± 0.04	0.18 ± 0.02
Adv. Deb.	0.71 ± 0.00	0.42 ± 0.00	0.60 ± 0.01	0.28 ± 0.00	0.86 ± 0.00	0.60 ± 0.01	1.47 ± 0.01	0.00 ± 0.00	0.88 ± 0.02
xFAIR	0.83 ± 0.00	0.44 ± 0.00	0.83 ± 0.00	0.50 ± 0.00	0.39 ± 0.00	0.07 ± 0.00	0.12 ± 0.01	0.47 ± 0.06	0.15 ± 0.00
MLP_{F1}	0.85 ± 0.00	0.55 ± 0.00	0.81 ± 0.00	0.46 ± 0.00	0.68 ± 0.00	0.11 ± 0.01	0.15 ± 0.02	0.67 ± 0.10	0.29 ± 0.02
MLP_{SP}	0.50 ± 0.00	0.20 ± 0.04	0.37 ± 0.10	0.12 ± 0.02	0.70 ± 0.14	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.06 ± 0.00
MLP_{EDiff}	0.51 ± 0.03	0.18 ± 0.05	0.51 ± 0.10	0.12 ± 0.03	0.53 ± 0.15	0.04 ± 0.02	0.10 ± 0.05	0.00 ± 0.00	0.07 ± 0.00
MLP_{F1+SP}	0.84 ± 0.00	0.55 ± 0.00	0.81 ± 0.00	0.45 ± 0.00	0.69 ± 0.01	0.12 ± 0.01	0.17 ± 0.02	0.00 ± 0.00	0.34 ± 0.02
$MLP_{F1+EDiff}$	0.84 ± 0.00	0.53 ± 0.00	0.80 ± 0.00	0.43 ± 0.00	0.68 ± 0.00	0.11 ± 0.00	0.16 ± 0.02	0.23 ± 0.10	0.16 ± 0.01
$MLP_{SP+EDiff}$	0.50 ± 0.00	0.15 ± 0.04	0.50 ± 0.10	0.10 ± 0.03	0.50 ± 0.16	0.00 ± 0.00	0.01 ± 0.01	0.00 ± 0.00	0.06 ± 0.00
MLP_{All}	0.84 ± 0.00	0.52 ± 0.00	0.80 ± 0.00	0.43 ± 0.00	0.67 ± 0.00	0.05 ± 0.00	0.08 ± 0.01	0.01 ± 0.00	0.15 ± 0.01
$MLP_{Curriculum}$	0.84 ± 0.00	0.52 ± 0.00	0.80 ± 0.00	0.43 ± 0.00	0.67 ± 0.00	0.05 ± 0.00	0.08 ± 0.01	0.01 ± 0.00	0.15 ± 0.01

Table A3: Detailed results for **DUTCH CENSUS - SEX 2**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.73 ± 0.03	0.79 ± 0.01	0.74 ± 0.02	0.70 ± 0.03	0.92 ± 0.01	0.13 ± 0.01	0.06 ± 0.01	0.10 ± 0.02	0.41 ± 0.05
Hardt	0.79 ± 0.00	0.82 ± 0.00	0.80 ± 0.00	0.77 ± 0.00	0.88 ± 0.00	0.17 ± 0.00	0.02 ± 0.00	0.03 ± 0.01	0.62 ± 0.01
Feldman	0.90 ± 0.00	0.85 ± 0.00	0.84 ± 0.00	0.82 ± 0.00	0.89 ± 0.00	0.30 ± 0.00	0.21 ± 0.00	0.10 ± 0.02	0.44 ± 0.03
Adv. Deb.	0.82 ± 0.00	0.83 ± 0.00	0.82 ± 0.00	0.80 ± 0.00	0.88 ± 0.00	0.20 ± 0.01	0.09 ± 0.00	0.00 ± 0.00	0.87 ± 0.02
xFAIR	0.87 ± 0.00	0.81 ± 0.00	0.80 ± 0.00	0.81 ± 0.00	0.81 ± 0.00	0.41 ± 0.00	0.58 ± 0.01	0.00 ± 0.00	0.58 ± 0.01
MLP_{F1}	0.90 ± 0.00	0.85 ± 0.00	0.84 ± 0.00	0.82 ± 0.00	0.89 ± 0.00	0.30 ± 0.00	0.22 ± 0.00	0.00 ± 0.00	0.60 ± 0.09
MLP_{SP}	0.48 ± 0.01	0.48 ± 0.10	0.51 ± 0.01	0.37 ± 0.08	0.70 ± 0.14	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00
MLP_{EDiff}	0.46 ± 0.03	0.28 ± 0.10	0.49 ± 0.01	0.36 ± 0.07	0.35 ± 0.14	0.03 ± 0.01	0.05 ± 0.02	0.06 ± 0.04	0.02 ± 0.00
MLP_{F1+SP}	0.81 ± 0.01	0.85 ± 0.00	0.83 ± 0.00	0.79 ± 0.01	0.91 ± 0.01	0.22 ± 0.01	0.11 ± 0.01	0.02 ± 0.01	0.24 ± 0.02
$MLP_{F1+EDiff}$	0.90 ± 0.00	0.84 ± 0.00	0.82 ± 0.00	0.79 ± 0.00	0.91 ± 0.00	0.22 ± 0.01	0.12 ± 0.01	0.00 ± 0.00	0.08 ± 0.01
$MLP_{SP+EDiff}$	0.47 ± 0.02	0.21 ± 0.10	0.49 ± 0.01	0.19 ± 0.08	0.30 ± 0.14	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.02	0.01 ± 0.00
MLP_{All}	0.80 ± 0.00	0.83 ± 0.00	0.80 ± 0.00	0.76 ± 0.00	0.92 ± 0.00	0.12 ± 0.00	0.14 ± 0.01	0.10 ± 0.07	0.05 ± 0.00
$MLP_{Curriculum}$	0.80 ± 0.00	0.83 ± 0.00	0.80 ± 0.00	0.76 ± 0.00	0.92 ± 0.00	0.12 ± 0.00	0.14 ± 0.01	0.08 ± 0.05	0.05 ± 0.00

Table A4: Detailed results for **GERMAN CREDIT - SEX**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.50 ± 0.00	0.82 ± 0.00	0.69 ± 0.01	0.69 ± 0.01	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.27 ± 0.00
Hardt	0.51 ± 0.01	0.07 ± 0.07	0.34 ± 0.03	0.08 ± 0.07	0.07 ± 0.07	0.01 ± 0.01	0.01 ± 0.01	0.00 ± 0.00	0.29 ± 0.01
Feldman	0.68 ± 0.02	0.81 ± 0.01	0.70 ± 0.01	0.72 ± 0.01	0.94 ± 0.02	0.08 ± 0.03	0.15 ± 0.06	0.00 ± 0.00	0.29 ± 0.01
Adv. Deb.	0.47 ± 0.01	0.77 ± 0.01	0.64 ± 0.01	0.68 ± 0.01	0.90 ± 0.01	0.03 ± 0.01	0.07 ± 0.01	0.00 ± 0.00	0.33 ± 0.00
xFAIR	0.78 ± 0.01	0.83 ± 0.01	0.75 ± 0.01	0.77 ± 0.01	0.92 ± 0.01	0.08 ± 0.01	0.16 ± 0.03	0.00 ± 0.00	0.32 ± 0.00
MLP_{F1}	0.69 ± 0.02	0.82 ± 0.01	0.70 ± 0.01	0.72 ± 0.01	0.95 ± 0.02	0.09 ± 0.03	0.16 ± 0.07	0.00 ± 0.00	0.29 ± 0.01
MLP_{SP}	0.51 ± 0.02	0.57 ± 0.12	0.57 ± 0.06	0.48 ± 0.10	0.70 ± 0.14	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.28 ± 0.00
MLP_{EDiff}	0.49 ± 0.02	0.45 ± 0.11	0.50 ± 0.05	0.48 ± 0.10	0.49 ± 0.13	0.01 ± 0.00	0.04 ± 0.02	0.00 ± 0.00	0.28 ± 0.00
MLP_{F1+SP}	0.67 ± 0.02	0.82 ± 0.00	0.71 ± 0.01	0.72 ± 0.01	0.95 ± 0.02	0.03 ± 0.02	0.08 ± 0.04	0.00 ± 0.00	0.28 ± 0.00
$MLP_{F1+EDiff}$	0.57 ± 0.01	0.82 ± 0.00	0.69 ± 0.01	0.69 ± 0.01	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.27 ± 0.00
$MLP_{SP+EDiff}$	0.50 ± 0.02	0.48 ± 0.12	0.53 ± 0.06	0.49 ± 0.10	0.55 ± 0.15	0.01 ± 0.01	0.04 ± 0.03	0.00 ± 0.00	0.27 ± 0.00
MLP_{All}	0.57 ± 0.01	0.82 ± 0.00	0.69 ± 0.01	0.69 ± 0.01	1.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.27 ± 0.00
$MLP_{Curriculum}$	0.57 ± 0.01	0.82 ± 0.00	0.69 ± 0.01	0.69 ± 0.01	1.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.27 ± 0.00

Table A5: Detailed results for **BANK MARKETING - MARITAL MARRIED**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.68 ± 0.01	0.43 ± 0.01	0.88 ± 0.00	0.47 ± 0.01	0.41 ± 0.03	0.03 ± 0.00	0.04 ± 0.00	0.60 ± 0.10	0.19 ± 0.02
Hardt	0.52 ± 0.01	0.09 ± 0.03	0.62 ± 0.11	0.05 ± 0.02	0.39 ± 0.15	0.01 ± 0.00	0.09 ± 0.06	0.09 ± 0.04	
Feldman	0.90 ± 0.00	0.58 ± 0.00	0.89 ± 0.00	0.53 ± 0.01	0.63 ± 0.01	0.01 ± 0.00	0.11 ± 0.01	0.60 ± 0.10	0.12 ± 0.01
Adv. Deb.	0.58 ± 0.00	0.26 ± 0.00	0.84 ± 0.00	0.29 ± 0.01	0.25 ± 0.01	0.16 ± 0.01	0.56 ± 0.02	0.00 ± 0.00	1.27 ± 0.02
xFAIR	0.93 ± 0.00	0.50 ± 0.00	0.91 ± 0.00	0.65 ± 0.01	0.41 ± 0.01	0.04 ± 0.00	0.10 ± 0.01	0.49 ± 0.09	0.11 ± 0.00
MLP_{F1}	0.90 ± 0.00	0.58 ± 0.00	0.89 ± 0.00	0.52 ± 0.01	0.67 ± 0.00	0.05 ± 0.00	0.06 ± 0.01	0.24 ± 0.08	0.14 ± 0.01
MLP_{SP}	0.49 ± 0.01	0.15 ± 0.03	0.35 ± 0.11	0.09 ± 0.01	0.70 ± 0.14	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01	0.02 ± 0.00
MLP_{EDiff}	0.47 ± 0.03	0.14 ± 0.03	0.37 ± 0.11	0.09 ± 0.01	0.63 ± 0.13	0.05 ± 0.02	0.12 ± 0.05	0.01 ± 0.01	0.04 ± 0.00
MLP_{F1+SP}	0.90 ± 0.00	0.59 ± 0.00	0.89 ± 0.00	0.52 ± 0.01	0.68 ± 0.01	0.04 ± 0.00	0.04 ± 0.01	0.51 ± 0.10	0.12 ± 0.01
$MLP_{F1+EDiff}$	0.80 ± 0.00	0.47 ± 0.01	0.87 ± 0.00	0.46 ± 0.01	0.48 ± 0.01	0.05 ± 0.00	0.08 ± 0.01	0.47 ± 0.12	0.06 ± 0.00
$MLP_{SP+EDiff}$	0.50 ± 0.01	0.14 ± 0.03	0.38 ± 0.11	0.08 ± 0.02	0.64 ± 0.14	0.00 ± 0.00	0.02 ± 0.01	0.01 ± 0.01	0.02 ± 0.00
MLP_{All}	0.79 ± 0.00	0.46 ± 0.01	0.88 ± 0.00	0.46 ± 0.01	0.46 ± 0.01	0.03 ± 0.00	0.04 ± 0.01	0.61 ± 0.13	0.05 ± 0.00
$MLP_{Curriculum}$	0.79 ± 0.00	0.46 ± 0.01	0.88 ± 0.00	0.46 ± 0.01	0.46 ± 0.01	0.03 ± 0.00	0.04 ± 0.01	0.62 ± 0.13	0.05 ± 0.00

Table A6: Detailed results for **CREDIT CARD CLIENTS - SEX**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.69 ± 0.00	0.52 ± 0.01	0.78 ± 0.01	0.51 ± 0.01	0.52 ± 0.01	0.02 ± 0.00	0.04 ± 0.01	0.20 ± 0.12	0.48 ± 0.05
Hardt	0.52 ± 0.02	0.12 ± 0.06	0.67 ± 0.06	0.08 ± 0.04	0.27 ± 0.13	0.00 ± 0.00	0.01 ± 0.00	0.29 ± 0.09	0.18 ± 0.09
Feldman	0.76 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.52 ± 0.00	0.55 ± 0.00	0.04 ± 0.00	0.06 ± 0.01	0.20 ± 0.12	0.25 ± 0.03
Adv. Deb.	0.59 ± 0.00	0.32 ± 0.01	0.78 ± 0.01	0.61 ± 0.04	0.24 ± 0.03	0.11 ± 0.04	0.31 ± 0.10	0.00 ± 0.00	1.16 ± 0.10
xFAIR	0.76 ± 0.00	0.47 ± 0.00	0.81 ± 0.00	0.64 ± 0.00	0.38 ± 0.00	0.02 ± 0.00	0.05 ± 0.01	0.45 ± 0.05	0.14 ± 0.00
MLP_{F1}	0.75 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.52 ± 0.00	0.55 ± 0.01	0.04 ± 0.00	0.06 ± 0.01	0.25 ± 0.09	0.30 ± 0.03
MLP_{SP}	0.49 ± 0.01	0.36 ± 0.00	0.22 ± 0.00	0.22 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.12 ± 0.03	0.04 ± 0.00
MLP_{EDiff}	0.47 ± 0.02	0.32 ± 0.03	0.31 ± 0.05	0.21 ± 0.01	0.83 ± 0.10	0.02 ± 0.01	0.05 ± 0.02	0.18 ± 0.09	0.05 ± 0.00
MLP_{F1+SP}	0.75 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.52 ± 0.00	0.54 ± 0.01	0.03 ± 0.00	0.04 ± 0.00	0.91 ± 0.06	0.12 ± 0.01
$MLP_{F1+EDiff}$	0.75 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.53 ± 0.00	0.53 ± 0.00	0.04 ± 0.00	0.06 ± 0.01	0.71 ± 0.07	0.08 ± 0.00
$MLP_{SP+EDiff}$	0.51 ± 0.01	0.33 ± 0.03	0.28 ± 0.05	0.20 ± 0.02	0.90 ± 0.09	0.00 ± 0.00	0.00 ± 0.00	0.14 ± 0.07	0.04 ± 0.00
MLP_{All}	0.75 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.53 ± 0.00	0.53 ± 0.01	0.03 ± 0.00	0.04 ± 0.01	0.97 ± 0.01	0.07 ± 0.00
$MLP_{Curriculum}$	0.74 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.53 ± 0.00	0.53 ± 0.01	0.03 ± 0.00	0.04 ± 0.01	0.95 ± 0.01	0.07 ± 0.00

Table A7: Detailed results for **CREDIT CARD CLIENTS - MARRIAGE 2**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.70 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.53 ± 0.00	0.54 ± 0.00	0.02 ± 0.00	0.03 ± 0.01	0.91 ± 0.03	0.29 ± 0.01
Hardt	0.53 ± 0.02	0.16 ± 0.06	0.62 ± 0.07	0.10 ± 0.04	0.37 ± 0.14	0.00 ± 0.00	0.01 ± 0.01	0.02 ± 0.01	0.30 ± 0.07
Feldman	0.75 ± 0.00	0.54 ± 0.00	0.79 ± 0.00	0.52 ± 0.00	0.55 ± 0.00	0.02 ± 0.00	0.04 ± 0.01	0.91 ± 0.03	0.18 ± 0.03
Adv. Deb.	0.60 ± 0.00	0.34 ± 0.00	0.80 ± 0.00	0.68 ± 0.01	0.22 ± 0.00	0.01 ± 0.00	0.03 ± 0.00	0.81 ± 0.05	0.22 ± 0.01
xFAIR	0.76 ± 0.00	0.47 ± 0.00	0.81 ± 0.00	0.64 ± 0.00	0.38 ± 0.01	0.01 ± 0.00	0.03 ± 0.00	0.18 ± 0.03	0.17 ± 0.00
MLP_{F1}	0.75 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.52 ± 0.00	0.55 ± 0.00	0.02 ± 0.00	0.03 ± 0.00	0.52 ± 0.13	0.29 ± 0.01
MLP_{SP}	0.52 ± 0.01	0.29 ± 0.05	0.34 ± 0.07	0.18 ± 0.03	0.80 ± 0.13	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.21 ± 0.01
MLP_{EDiff}	0.54 ± 0.02	0.24 ± 0.05	0.47 ± 0.07	0.15 ± 0.03	0.56 ± 0.13	0.07 ± 0.04	0.13 ± 0.08	0.03 ± 0.03	0.24 ± 0.01
MLP_{F1+SP}	0.75 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.52 ± 0.00	0.55 ± 0.00	0.01 ± 0.00	0.03 ± 0.00	0.74 ± 0.11	0.26 ± 0.01
$MLP_{F1+EDiff}$	0.75 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.53 ± 0.00	0.53 ± 0.01	0.02 ± 0.00	0.03 ± 0.00	0.26 ± 0.11	0.27 ± 0.01
$MLP_{SP+EDiff}$	0.51 ± 0.02	0.19 ± 0.05	0.50 ± 0.09	0.17 ± 0.04	0.51 ± 0.16	0.00 ± 0.00	0.01 ± 0.00	0.05 ± 0.04	0.22 ± 0.01
MLP_{All}	0.74 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.53 ± 0.00	0.53 ± 0.01	0.01 ± 0.00	0.03 ± 0.00	0.69 ± 0.11	0.26 ± 0.01
$MLP_{Curriculum}$	0.74 ± 0.00	0.53 ± 0.00	0.79 ± 0.00	0.53 ± 0.00	0.53 ± 0.01	0.01 ± 0.00	0.03 ± 0.00	0.73 ± 0.10	0.25 ± 0.01

Table A8: Detailed results for **OULAD - GENDER**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.50 ± 0.00	0.81 ± 0.00	0.68 ± 0.00	0.68 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.91 ± 0.04	0.33 ± 0.01
Hardt	0.50 ± 0.00	0.00 ± 0.00	0.32 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.13 ± 0.04	0.03 ± 0.00
Feldman	0.51 ± 0.00	0.81 ± 0.00	0.68 ± 0.00	0.68 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.91 ± 0.04	0.09 ± 0.00
Adv. Deb.	0.51 ± 0.00	0.81 ± 0.00	0.68 ± 0.00	0.68 ± 0.00	0.98 ± 0.00	0.03 ± 0.00	0.08 ± 0.01	0.00 ± 0.00	0.19 ± 0.01
xFAIR	0.60 ± 0.00	0.75 ± 0.00	0.64 ± 0.00	0.71 ± 0.00	0.81 ± 0.00	0.02 ± 0.00	0.04 ± 0.01	0.04 ± 0.02	0.27 ± 0.01
MLP_{F1}	0.50 ± 0.00	0.81 ± 0.00	0.68 ± 0.00	0.68 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.08 ± 0.00
MLP_{SP}	0.51 ± 0.01	0.65 ± 0.10	0.60 ± 0.05	0.61 ± 0.06	0.80 ± 0.13	0.00 ± 0.00	0.00 ± 0.00	0.14 ± 0.06	0.03 ± 0.00
MLP_{EDiff}	0.49 ± 0.01	0.56 ± 0.10	0.55 ± 0.05	0.60 ± 0.06	0.65 ± 0.14	0.04 ± 0.03	0.08 ± 0.06	0.01 ± 0.01	0.05 ± 0.00
MLP_{F1+SP}	0.50 ± 0.00	0.81 ± 0.00	0.68 ± 0.00	0.68 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.06 ± 0.00
$MLP_{F1+EDiff}$	0.50 ± 0.01	0.81 ± 0.00	0.68 ± 0.00	0.68 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.09 ± 0.08	0.06 ± 0.01
$MLP_{SP+EDiff}$	0.50 ± 0.01	0.51 ± 0.12	0.54 ± 0.05	0.54 ± 0.09	0.61 ± 0.15	0.00 ± 0.00	0.01 ± 0.00	0.16 ± 0.08	0.04 ± 0.00
MLP_{All}	0.51 ± 0.01	0.81 ± 0.00	0.68 ± 0.00	0.68 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.30 ± 0.10	0.05 ± 0.00
$MLP_{Curriculum}$	0.51 ± 0.01	0.81 ± 0.00	0.68 ± 0.00	0.68 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.20 ± 0.06	0.05 ± 0.00

Table A9: Detailed results for **LAWSCHOOL - MALE**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.50 ± 0.00	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.16 ± 0.06	0.18 ± 0.00
Hardt	0.74 ± 0.02	0.89 ± 0.00	0.82 ± 0.01	0.95 ± 0.00	0.84 ± 0.01	0.02 ± 0.00	0.05 ± 0.01	0.50 ± 0.08	0.15 ± 0.02
Feldman	0.86 ± 0.00	0.95 ± 0.00	0.90 ± 0.00	0.91 ± 0.00	0.99 ± 0.00	0.02 ± 0.00	0.06 ± 0.01	0.16 ± 0.06	0.08 ± 0.01
Adv. Deb.	0.50 ± 0.00	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.42 ± 0.01
xFAIR	0.85 ± 0.00	0.94 ± 0.00	0.90 ± 0.00	0.91 ± 0.00	0.97 ± 0.00	0.02 ± 0.00	0.11 ± 0.01	0.37 ± 0.04	0.08 ± 0.00
MLP_{F1}	0.85 ± 0.01	0.95 ± 0.00	0.90 ± 0.00	0.91 ± 0.00	0.99 ± 0.00	0.02 ± 0.00	0.06 ± 0.01	0.33 ± 0.11	0.13 ± 0.01
MLP_{SP}	0.51 ± 0.02	0.28 ± 0.14	0.34 ± 0.11	0.27 ± 0.13	0.30 ± 0.14	0.00 ± 0.00	0.00 ± 0.00	0.04 ± 0.02	0.02 ± 0.00
MLP_{EDiff}	0.49 ± 0.03	0.18 ± 0.10	0.23 ± 0.08	0.34 ± 0.13	0.16 ± 0.10	0.06 ± 0.04	0.13 ± 0.08	0.00 ± 0.00	0.03 ± 0.00
MLP_{F1+SP}	0.81 ± 0.03	0.94 ± 0.00	0.89 ± 0.00	0.90 ± 0.00	0.99 ± 0.00	0.01 ± 0.00	0.02 ± 0.01	0.41 ± 0.12	0.05 ± 0.01
$MLP_{F1+EDiff}$	0.76 ± 0.02	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.02 ± 0.01	0.05 ± 0.01
$MLP_{SP+EDiff}$	0.47 ± 0.03	0.12 ± 0.09	0.20 ± 0.07	0.18 ± 0.12	0.11 ± 0.09	0.00 ± 0.00	0.00 ± 0.00	0.03 ± 0.03	0.02 ± 0.00
MLP_{All}	0.77 ± 0.02	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.56 ± 0.13	0.03 ± 0.00
$MLP_{Curriculum}$	0.77 ± 0.02	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.56 ± 0.12	0.03 ± 0.00

Table A10: Detailed results for **LAWSCHOOL - RACE**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.50 ± 0.00	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.56 ± 0.01
Hardt	0.64 ± 0.01	0.75 ± 0.02	0.64 ± 0.03	0.94 ± 0.00	0.64 ± 0.04	0.07 ± 0.01	0.06 ± 0.01	0.00 ± 0.00	0.40 ± 0.04
Feldman	0.86 ± 0.00	0.95 ± 0.00	0.90 ± 0.00	0.91 ± 0.00	0.99 ± 0.00	0.19 ± 0.01	0.52 ± 0.02	0.00 ± 0.00	0.17 ± 0.01
Adv. Deb.	0.50 ± 0.00	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.01	0.03 ± 0.00
xFAIR	0.85 ± 0.00	0.94 ± 0.00	0.90 ± 0.00	0.91 ± 0.00	0.97 ± 0.00	0.15 ± 0.00	0.38 ± 0.01	0.00 ± 0.00	0.13 ± 0.00
MLP_{F1}	0.85 ± 0.00	0.94 ± 0.00	0.90 ± 0.00	0.91 ± 0.00	0.98 ± 0.00	0.19 ± 0.01	0.51 ± 0.02	0.01 ± 0.01	0.26 ± 0.03
MLP_{SP}	0.53 ± 0.03	0.57 ± 0.15	0.58 ± 0.12	0.63 ± 0.13	0.60 ± 0.15	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01	0.03 ± 0.00
MLP_{EDiff}	0.47 ± 0.04	0.33 ± 0.13	0.36 ± 0.11	0.60 ± 0.13	0.33 ± 0.14	0.05 ± 0.03	0.10 ± 0.05	0.00 ± 0.00	0.04 ± 0.00
MLP_{F1+SP}	0.79 ± 0.01	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.08 ± 0.07	0.05 ± 0.00
$MLP_{F1+EDiff}$	0.74 ± 0.02	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.01 ± 0.01	0.04 ± 0.03	0.04 ± 0.03	0.05 ± 0.01
$MLP_{SP+EDiff}$	0.51 ± 0.02	0.42 ± 0.14	0.44 ± 0.12	0.63 ± 0.13	0.43 ± 0.15	0.00 ± 0.00	0.01 ± 0.01	0.05 ± 0.03	0.03 ± 0.00
MLP_{All}	0.70 ± 0.02	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01	0.00 ± 0.00	0.04 ± 0.00
$MLP_{Curriculum}$	0.70 ± 0.02	0.94 ± 0.00	0.89 ± 0.00	0.89 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.01	0.00 ± 0.00	0.04 ± 0.00

Table A11: Detailed results for **COMPAS - RACE CAUCASIAN**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.50 ± 0.00	0.65 ± 0.00	0.49 ± 0.00	0.49 ± 0.00	1.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.48 ± 0.06
Hardt	0.50 ± 0.00	0.00 ± 0.00	0.51 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.02 ± 0.01	0.05 ± 0.00
Feldman	0.72 ± 0.00	0.69 ± 0.01	0.64 ± 0.01	0.59 ± 0.01	0.84 ± 0.01	0.19 ± 0.01	0.32 ± 0.01	0.00 ± 0.00	0.12 ± 0.01
Adv. Deb.	0.41 ± 0.00	0.46 ± 0.00	0.41 ± 0.00	0.41 ± 0.00	0.52 ± 0.01	0.27 ± 0.01	0.52 ± 0.01	0.00 ± 0.00	0.65 ± 0.04
xFAIR	0.68 ± 0.00	0.63 ± 0.00	0.63 ± 0.00	0.62 ± 0.01	0.63 ± 0.01	0.18 ± 0.01	0.30 ± 0.01	0.01 ± 0.00	0.24 ± 0.01
MLP_{F1}	0.69 ± 0.03	0.69 ± 0.01	0.63 ± 0.02	0.59 ± 0.01	0.84 ± 0.02	0.16 ± 0.02	0.26 ± 0.03	0.51 ± 0.11	0.12 ± 0.01
MLP_{SP}	0.49 ± 0.01	0.26 ± 0.10	0.51 ± 0.01	0.25 ± 0.08	0.40 ± 0.15	0.00 ± 0.00	0.01 ± 0.01	0.06 ± 0.04	0.06 ± 0.00
MLP_{EDiff}	0.46 ± 0.02	0.22 ± 0.08	0.50 ± 0.01	0.32 ± 0.08	0.26 ± 0.11	0.12 ± 0.05	0.24 ± 0.10	0.00 ± 0.00	0.07 ± 0.00
MLP_{F1+SP}	0.64 ± 0.03	0.67 ± 0.01	0.57 ± 0.02	0.54 ± 0.02	0.90 ± 0.03	0.06 ± 0.02	0.10 ± 0.03	0.05 ± 0.03	0.11 ± 0.02
$MLP_{F1+EDiff}$	0.52 ± 0.04	0.66 ± 0.00	0.52 ± 0.01	0.51 ± 0.01	0.95 ± 0.02	0.04 ± 0.02	0.08 ± 0.03	0.07 ± 0.06	0.09 ± 0.01
$MLP_{SP+EDiff}$	0.48 ± 0.02	0.22 ± 0.09	0.50 ± 0.01	0.19 ± 0.07	0.32 ± 0.14	0.01 ± 0.01	0.02 ± 0.02	0.04 ± 0.02	0.06 ± 0.00
MLP_{All}	0.53 ± 0.02	0.66 ± 0.00	0.50 ± 0.01	0.49 ± 0.01	0.98 ± 0.01	0.01 ± 0.01	0.03 ± 0.02	0.22 ± 0.10	0.06 ± 0.00
$MLP_{Curriculum}$	0.53 ± 0.02	0.66 ± 0.00	0.50 ± 0.01	0.49 ± 0.01	0.98 ± 0.01	0.01 ± 0.01	0.03 ± 0.02	0.22 ± 0.10	0.06 ± 0.00

Table A12: Detailed results for **DIABETES - GENDER**

	AUC	F1	Accuracy	Precision	Recall	Statistical Parity	Equalized Odds	GPF_{FAE}	Explanation Difference
Agarwal	0.54 ± 0.02	0.23 ± 0.05	0.71 ± 0.02	0.30 ± 0.07	0.22 ± 0.05	0.09 ± 0.02	0.31 ± 0.07	0.10 ± 0.04	0.64 ± 0.01
Hardt	0.47 ± 0.04	0.30 ± 0.04	0.42 ± 0.04	0.21 ± 0.03	0.57 ± 0.05	0.12 ± 0.02	0.46 ± 0.08	0.19 ± 0.06	0.67 ± 0.01
Feldman	0.49 ± 0.04	0.29 ± 0.03	0.38 ± 0.07	0.21 ± 0.02	0.66 ± 0.10	0.08 ± 0.02	0.30 ± 0.07	0.10 ± 0.04	0.60 ± 0.02
Adv. Deb.	0.54 ± 0.02	0.36 ± 0.03	0.38 ± 0.02	0.23 ± 0.02	0.82 ± 0.03	0.15 ± 0.04	0.41 ± 0.05	0.09 ± 0.03	0.74 ± 0.01
xFAIR	0.64 ± 0.02	0.00 ± 0.00	0.79 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.12 ± 0.05	0.67 ± 0.02
MLP_{F1}	0.53 ± 0.04	0.34 ± 0.02	0.33 ± 0.05	0.24 ± 0.03	0.86 ± 0.04	0.04 ± 0.02	0.17 ± 0.08	0.10 ± 0.03	0.58 ± 0.02
MLP_{SP}	0.52 ± 0.03	0.27 ± 0.05	0.42 ± 0.08	0.17 ± 0.03	0.67 ± 0.13	0.01 ± 0.01	0.10 ± 0.08	0.11 ± 0.05	0.57 ± 0.01
MLP_{EDiff}	0.52 ± 0.03	0.24 ± 0.06	0.42 ± 0.08	0.15 ± 0.04	0.64 ± 0.14	0.00 ± 0.00	0.08 ± 0.07	0.09 ± 0.04	0.57 ± 0.02
MLP_{F1+SP}	0.54 ± 0.05	0.35 ± 0.02	0.37 ± 0.05	0.25 ± 0.03	0.83 ± 0.07	0.07 ± 0.02	0.20 ± 0.06	0.09 ± 0.03	0.57 ± 0.01
$MLP_{F1+EDiff}$	0.53 ± 0.04	0.34 ± 0.02	0.33 ± 0.05	0.24 ± 0.03	0.86 ± 0.08	0.04 ± 0.02	0.17 ± 0.08	0.10 ± 0.05	0.58 ± 0.01
$MLP_{SP+EDiff}$	0.53 ± 0.03	0.24 ± 0.06	0.42 ± 0.08	0.15 ± 0.04	0.64 ± 0.14	0.00 ± 0.00	0.08 ± 0.07	0.12 ± 0.05	0.55 ± 0.01
MLP_{All}	0.56 ± 0.04	0.34 ± 0.03	0.28 ± 0.04	0.22 ± 0.02	0.90 ± 0.05	0.05 ± 0.02	0.20 ± 0.08	0.11 ± 0.04	0.61 ± 0.02
$MLP_{Curriculum}$	0.56 ± 0.04	0.34 ± 0.03	0.28 ± 0.04	0.22 ± 0.02	0.90 ± 0.05	0.05 ± 0.02	0.20 ± 0.08	0.10 ± 0.03	0.58 ± 0.01