Table S2. Fairness Metrics

Metric Name	General Definition	Example Statement	Equality Statement	Formula
Independence (Demographic Parity)	A classifier (M) satisfies the independence fairness criterion, if its output ($\hat{Y}$) is independent of the sensitive attributes *A*.	The model's prediction of whether a glaucoma patient will progress to surgery is independent of that patient's sensitive characteristics; the proportion of those in the positive predicted class is the same across sensitive characteristics.	$P\{\hat{Y} = 1 \mid A = a_1\} = P\{\hat{Y} = 1 \mid A = a_2\}$	|Support(M,G1) - Support(M, G2)|
Equalized Odds	A classifier (M) satisfies the equalized odds definition of fairness, if its output ($\hat{Y}$) is conditionally (with respect to the target output *Y*) independent of the sensitive attribute *A*.	The model's prediction of whether a glaucoma patient will progress to surgery performs with equal false positive and true positive rates for White and nonWhite patients.	$P\{\hat{Y} = 1 \mid Y = 1, A = a_1\} = P\{\hat{Y} = 1 \mid Y = 1, \quad A = a_2\}$ $P\{\hat{Y} = 1 \mid Y = 0, A = a_1\} = P\{\hat{Y} = 1 \mid Y = 0, \quad A = a_2\}$	|Recall(M,G1) - Recall(M, G2)| + |FPR(M,G1) - FPR(M, G2)|
Overall Accuracy Equality	A classifier (M) satisfies the overall accuracy equality, if a learning system's accuracy values for both groups are equal.	The model's prediction of whether a glaucoma patient will progress to surgery is equally accurate for White and Non-White patients	$P\{\hat{Y} = Y \mid A = a_1\} = P\{\hat{Y} = Y \mid A = a_2\}$	|Accuracy(M,G1) - Accuracy(M, G2)|
Sufficiency (Calibration)	A classifier (M) satisfies the sufficiency condition if the target output *Y* is conditionally (with respect to $\hat{Y}$) independent of the sensitive attribute *A*.	For any given predicted probability score, the actual probability of whether a glaucoma patient will progress to surgery should be equal for both White and Non-White patients.	$P\{Y = 1 \mid \hat{Y} = \hat{y}, A = a_1\} = P\{Y = 1 \mid \hat{Y} = \hat{y}, \quad A = a_2\}$	|Precision(M,G1) - Precision(M, G2)| + |FOR(M,G1) - FOR(M, G2)|

Let $A_1 \ldots, A_M$ be the sensitive attributes and $X_1, \ldots, X_m$be the other (insensitive attributes). Y is the target (desired) output and $\hat{Y}$ is the classifier output. Furthermore, assume A has two possible values a_1 and a_2. Finally, we define $A = a_1$ as G1 and $A = a_2$ as G2.