

Data Collection Plan for Driver Monitoring System (DMS)

This document presents a detailed data collection matrix, aimed at building a diverse and robust dataset for training AI models. The plan focuses on diversity in demographics (nationality, gender, skin tone), behavioral states, and environmental conditions.

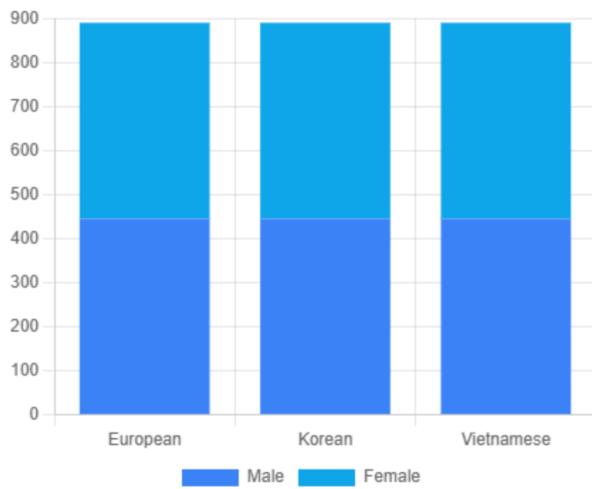
Data Collection Matrix (Target Number of Image/Video Clips)

The table below defines the number of data samples to be collected for each combination of behavior and environmental condition, segmented by nationality and gender. This segmentation ensures diversity in facial morphology, gender, and skin tone. The data has been balanced across states for more effective model training.

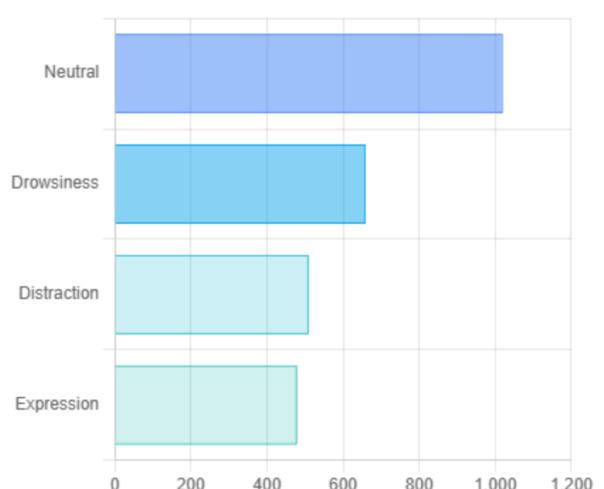
Category (State & Environment)	European		Korean		Vietnamese		Total
	Male	Female	Male	Female	Male	Female	
Focused Driving (Baseline)	50	50	50	50	50	50	300
Drowsiness (Yawn) - Good Light	20	20	20	20	20	20	120
Drowsiness (Yawn) - Low Light	30	30	30	30	30	30	180
Drowsiness (Long Eye Closure) - Good Light	25	25	25	25	25	25	150
Drowsiness (Long Eye Closure) - Low Light	35	35	35	35	35	35	210
Distraction (Phone Use) - Good Light	30	30	30	30	30	30	180
Distraction (Phone Use) - Low Light	30	30	30	30	30	30	180
Distraction (Looking Away) - Good Light	25	25	25	25	25	25	150
Expression (Smile) - Good Light	40	40	40	40	40	40	240
Expression (Talking) - Good Light	40	40	40	40	40	40	240
Occlusion (Glasses) - Good Light	40	40	40	40	40	40	240
Occlusion (Face Mask) - Good Light	40	40	40	40	40	40	240
Environment (Sun Glare) - No Occlusion	40	40	40	40	40	40	240
GRAND TOTAL	445	445	445	445	445	445	2,670

Data Collection Target Visualization

Data Distribution by Gender & Nationality



Data Distribution by Core State



Data Distribution by Environmental Condition

