Table of Contents

Engineered Features

Overview

The features are organized into **5 categories**:

Category	Count	Purpose
Edge-Based	5	Detect edges, lane markings, road boundaries
Spatial/Region	11	Capture where brightness is located (critical for steering)
Texture	7	Describe surface patterns using LBP and GLCM
Histogram/Distribution	11	Overall pixel intensity distribution

Category	Count	Purpose
Gradient/HOG	4	Shape and edge direction information

1. Edge-Based Features (5 features)

These features capture edge information using gradient operators, useful for detecting lane markings and road boundaries.

Feature	How It's Computed	What It Tells Us
edge_mean	Mean of Sobel edge magnitude ($\sqrt{Gx^2 + Gy^2}$)	Overall edge intensity in the image; higher = more edges/details
edge_std	Standard deviation of edge magnitude	Variation in edge strength; higher = more heterogeneous edges
edge_max	Maximum edge magnitude	Strongest edge in the image
edge_hv_ratio	Horizontal edge energy / Vertical edge energy	Ratio > 1 means more horizontal edges (e.g., road horizon); < 1 means more vertical (e.g., lane lines)
canny_density	Fraction of pixels detected as edges by Canny	Overall "edginess" of the image; low = smooth scene

2. Spatial/Region Features (11 features)

These capture **where** brightness is located in the image — critical for steering direction prediction.

Feature	How It's Computed	What It Tells Us
lr_asymmetry	Mean(left half) - Mean(right half)	**Key feature!** Positive = left side brighter (suggests left turn)

Feature	How It's Computed	What It Tells Us
lr_ratio	Mean(left half) / Mean(right half)	Similar to above, but as a ratio
tb_ratio	Mean(top half) / Mean(bottom half)	Sky vs road brightness; higher = brighter sky
tb_diff	Mean(top half) - Mean(bottom half)	Absolute brightness difference top vs bottom
quad_tl_mean	Mean intensity of top-left quadrant	Brightness in upper-left region
quad_tr_mean	Mean intensity of top-right quadrant	Brightness in upper-right region
quad_bl_mean	Mean intensity of bottom-left quadrant	Brightness in lower-left region
quad_br_mean	Mean intensity of bottom-right quadrant	Brightness in lower-right region
diag_asymmetry	(TL + BR) - (TR + BL)	Diagonal brightness pattern
brightness_center_x	Weighted centroid X (normalized to [-1, 1])	**Key feature!** Negative = brightness center is left; positive = right
brightness_center_y	Weighted centroid Y (normalized to [-1, 1])	Negative = brightness center is up; positive = down

3. Texture Features (7 features)

These describe the surface texture patterns using Local Binary Patterns (LBP) and Gray Level Co-occurrence Matrix (GLCM).

Local Binary Pattern (LBP)

LBP compares each pixel to its neighbors, encoding local texture as a binary pattern.

Feature	How It's Computed	What It Tells Us
lbp_mean	Mean of LBP image	Average local texture pattern code
lbp_std	Std deviation of LBP values	Texture pattern variation
lbp_uniformity	Sum of squared LBP histogram bins	Higher = more uniform texture (fewer distinct patterns)

Gray Level Co-occurrence Matrix (GLCM)

GLCM analyzes how often pairs of pixels with specific values occur at specific spatial relationships.

Feature	How It's Computed	What It Tells Us
glcm_contrast	GLCM contrast (averaged over 4 angles)	Local intensity variation; higher = rougher texture
glcm_correlation	GLCM correlation	How correlated a pixel is with its neighbors; higher = more structured patterns
glcm_energy	GLCM energy (angular second moment)	Texture uniformity; higher = more regular/repetitive
glcm_homogeneity	GLCM homogeneity	Closeness of distribution to diagonal; higher = smoother transitions

4. Histogram/Distribution Features (11 features)

These capture the overall pixel intensity distribution.

Feature	How It's Computed	What It Tells Us
intensity_mean	Mean pixel value	Overall image brightness
intensity_std	Std deviation of pixel values	Image contrast; higher = more variation
percentile_10	10th percentile of pixel values	Darkest regions (shadows)
percentile_25	25th percentile (Q1)	Lower brightness bound
percentile_50	Median pixel value	Central brightness
percentile_75	75th percentile (Q3)	Upper brightness bound
percentile_90	90th percentile	Brightest regions (highlights)
iqr	Q3 - Q1 (interquartile range)	Spread of middle 50% of intensities
skewness	Skewness of pixel distribution	Negative = tail toward dark; Positive = tail toward bright
kurtosis	Kurtosis of pixel distribution	Peakedness; high = concentrated around mean; low = flat distribution
intensity_range	Max - Min pixel value	Dynamic range of the image

5. Gradient/HOG Features (4 features)

Summary statistics from Histogram of Oriented Gradients (HOG), which captures shape and edge direction information.

HOG divides the image into cells, computes gradient orientations within each cell, and creates histograms of these orientations. The full HOG descriptor has 1764 values, so we summarize it with statistics.

Feature	How It's Computed	What It Tells Us
hog_mean	Mean of full HOG descriptor (1764 values)	Average gradient orientation strength
hog_std	Std deviation of HOG values	Variation in gradient orientations
hog_max	Maximum HOG bin value	Strongest gradient direction
hog_energy	Sum of squared HOG values	Total gradient energy; higher = more edges/structure

Most Discriminative Features for Steering

Based on ANOVA F-scores and Mutual Information analysis, the **most important features for predicting steering direction** are:

Top Features

1. lr_asymmetry — Left-right brightness difference is directly related to steering direction
2. brightness_center_x — Where the "center of mass" of brightness is located horizontally
3. quad_bl_mean / quad_br_mean — Bottom quadrant brightness patterns
4. lr_ratio — Ratio version of left-right asymmetry
5. edge_hv_ratio — Balance of horizontal vs vertical edges

Why These Features Work

These spatial features make intuitive sense for steering prediction:

- **Left turns**: The left side of the road typically shows different features (obstacles, curves, lane markings) than the right side, causing asymmetry in brightness.
- **Right turns**: The opposite pattern occurs.

- **Forward/Straight**: Both sides tend to be more balanced.

The brightness_center_x feature essentially asks: "Where is the visual 'weight' of the image?" If it's shifted left, the vehicle may need to turn left to center the road.

Logistic Regression

Raw Data:
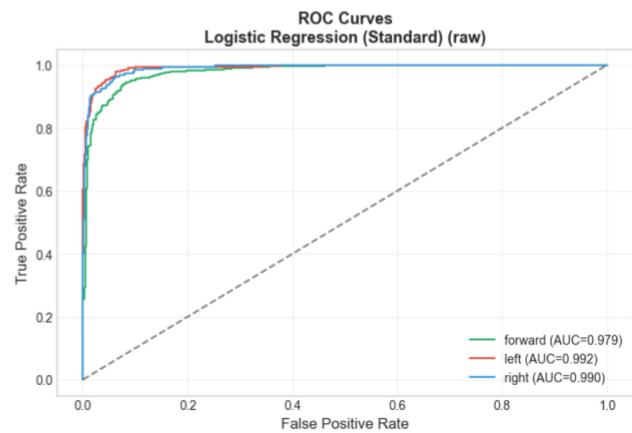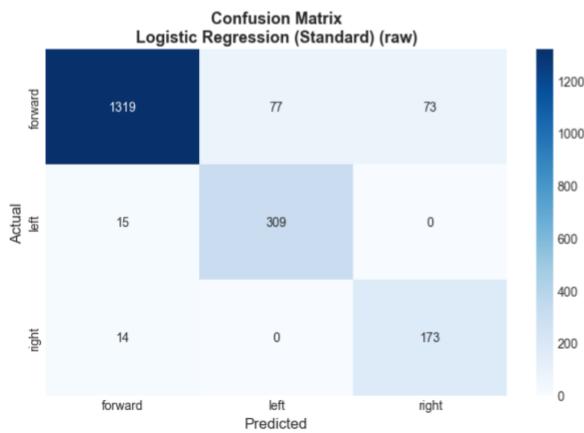
```
Training Standard Logistic Regression on RAW (PCA) features...

========================================================
Model: Logistic Regression (Standard) | Features: raw
========================================================
Accuracy:           0.9096
Balanced Accuracy: 0.9256
Macro F1:           0.8687
Weighted F1:        0.9127
Macro Precision:    0.8274
Macro Recall:       0.9256
ROC-AUC (macro):    0.9871

Per-class metrics:
  forward   : F1=0.9365, Precision=0.9785, Recall=0.8979
  left      : F1=0.8704, Precision=0.8005, Recall=0.9537
  right     : F1=0.7991, Precision=0.7033, Recall=0.9251

Classification Report:
              precision    recall  f1-score   support

     forward       0.98      0.90      0.94      1469
        left       0.80      0.95      0.87       324
       right       0.70      0.93      0.80       187

    accuracy                           0.91      1980
   macro avg       0.83      0.93      0.87      1980
weighted avg       0.92      0.91      0.91      1980
```


Confusion Matrix
Logistic Regression (Standard) (raw)

ROC Curves
Logistic Regression (Standard) (raw)

forward (AUC=0.979)
left (AUC=0.992)
right (AUC=0.990)

Engineered Features:

```
Training Standard Logistic Regression on ENGINEERED features...

========================================================
Model: Logistic Regression (Standard) | Features: engineered
========================================================
Accuracy:          0.8015
Balanced Accuracy: 0.8538
Macro F1:          0.7449
Weighted F1:       0.8130
Macro Precision:   0.6950
Macro Recall:      0.8538
ROC-AUC (macro):   0.9469

Per-class metrics:
  forward   : F1=0.8522, Precision=0.9521, Recall=0.7713
  left      : F1=0.7259, Precision=0.6225, Recall=0.8704
  right     : F1=0.6565, Precision=0.5104, Recall=0.9198

Classification Report:
              precision    recall  f1-score   support

     forward       0.95      0.77      0.85      1469
        left       0.62      0.87      0.73       324
       right       0.51      0.92      0.66       187

    accuracy                           0.80      1980
   macro avg       0.70      0.85      0.74      1980
weighted avg       0.86      0.80      0.81      1980
```
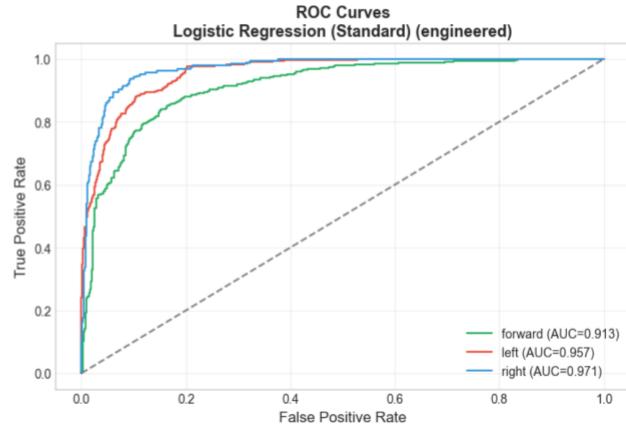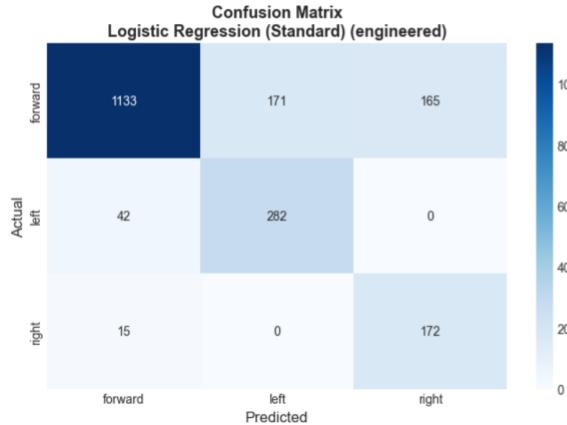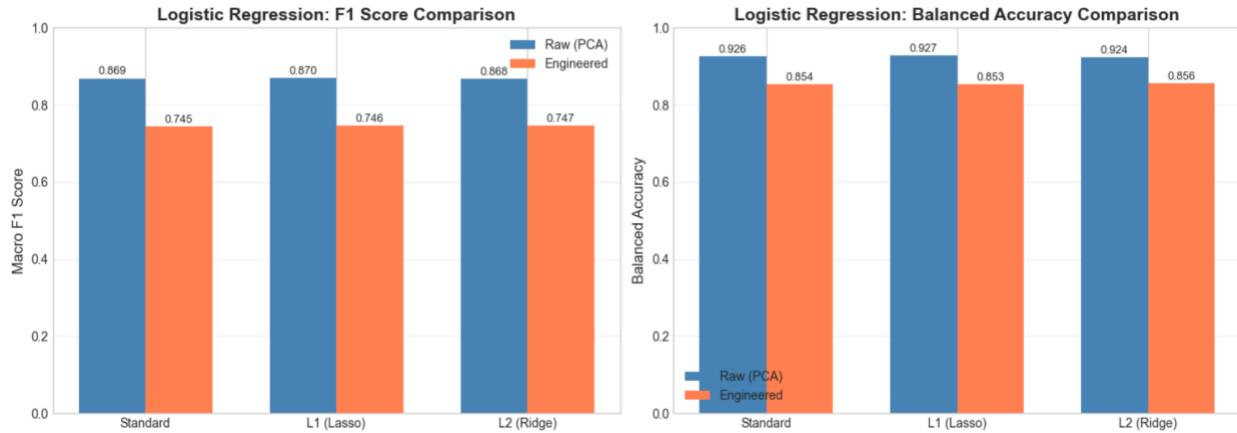

Confusion Matrix / ROC Curves — Logistic Regression (Standard) (engineered)

Summary:

```
================================================================================
LOGISTIC REGRESSION RESULTS SUMMARY
================================================================================
```

| Model	Features	Accuracy	Balanced Acc	F1 (Macro)	ROC-AUC
Logistic Regression (Standard)	raw	0.909596	0.925576	0.868652	0.987109
Logistic Regression (Standard)	engineered	0.801515	0.853810	0.744852	0.946896
Logistic Regression (L1/Lasso)	raw	0.910101	0.927358	0.869894	0.987304
Logistic Regression (L1/Lasso)	engineered	0.802525	0.853462	0.745890	0.946795
Logistic Regression (L2/Ridge)	raw	0.908586	0.924320	0.867686	0.987376
Logistic Regression (L2/Ridge)	engineered	0.803030	0.856046	0.746940	0.947006

Logistic Regression: F1 Score Comparison / Logistic Regression: Balanced Accuracy Comparison

KNN

Raw Data:
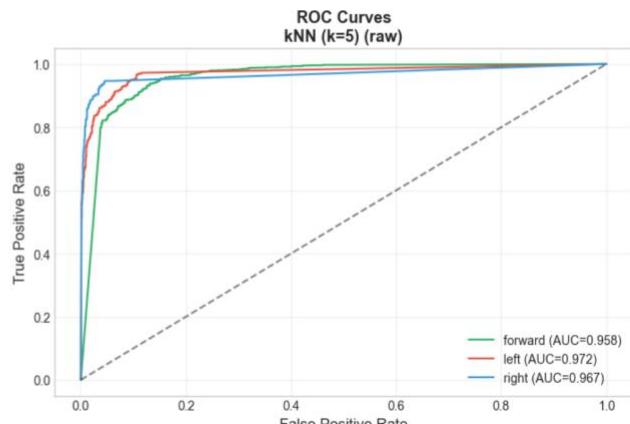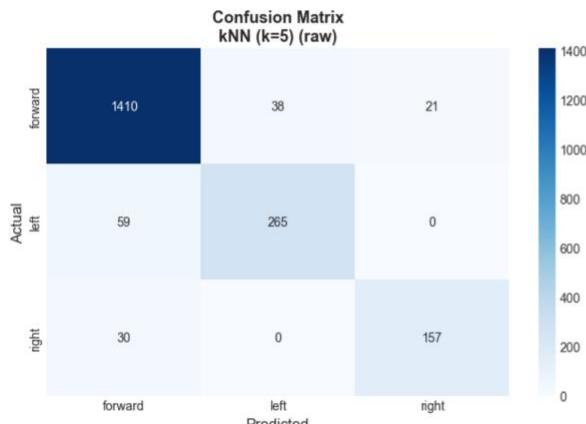
```
Training kNN (k=5) on RAW (PCA) features...

===========================================================
Model: kNN (k=5) | Features: raw
===========================================================
Accuracy:          0.9253
Balanced Accuracy: 0.8724
Macro F1:          0.8852
Weighted F1:       0.9245
Macro Precision:   0.8991
Macro Recall:      0.8724
ROC-AUC (macro):   0.9657

Per-class metrics:
  forward   : F1=0.9501, Precision=0.9406, Recall=0.9598
  left      : F1=0.8453, Precision=0.8746, Recall=0.8179
  right     : F1=0.8603, Precision=0.8820, Recall=0.8396

Classification Report:
              precision    recall  f1-score   support

     forward       0.94      0.96      0.95      1469
        left       0.87      0.82      0.85       324
       right       0.88      0.84      0.86       187

    accuracy                           0.93      1980
   macro avg       0.90      0.87      0.89      1980
weighted avg       0.92      0.93      0.92      1980
```
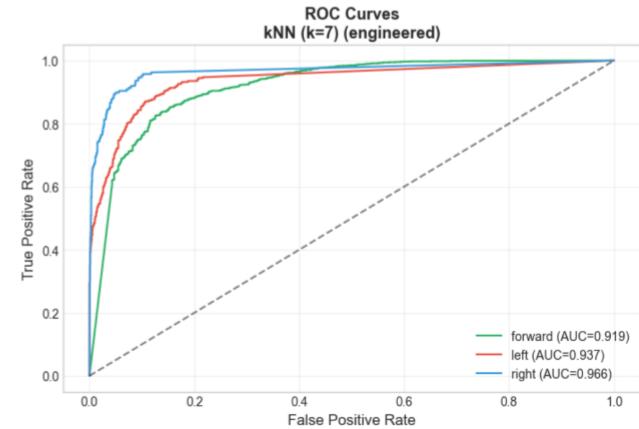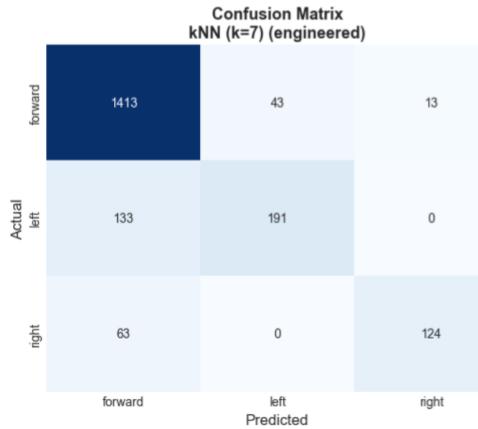
Engineered Features:

```
Training kNN (k=7) on ENGINEERED features...

========================================================
Model: kNN (k=7) | Features: engineered
========================================================
Accuracy:           0.8727
Balanced Accuracy:  0.7382
Macro F1:           0.7894
Weighted F1:        0.8655
Macro Precision:    0.8665
Macro Recall:       0.7382
ROC-AUC (macro):    0.9406

Per-class metrics:
  forward    : F1=0.9181, Precision=0.8782, Recall=0.9619
  left       : F1=0.6846, Precision=0.8162, Recall=0.5895
  right      : F1=0.7654, Precision=0.9051, Recall=0.6631

Classification Report:
              precision    recall  f1-score   support

     forward       0.88      0.96      0.92      1469
        left       0.82      0.59      0.68       324
       right       0.91      0.66      0.77       187

    accuracy                           0.87      1980
   macro avg       0.87      0.74      0.79      1980
weighted avg       0.87      0.87      0.87      1980
```


Confusion Matrix
kNN (k=7) (engineered)

ROC Curves
kNN (k=7) (engineered)

Naïve Bayes

Raw Data

```
Training Gaussian Naive Bayes on RAW (PCA) features...

============================================================
Model: Gaussian Naive Bayes | Features: raw
============================================================
Accuracy:          0.8702
Balanced Accuracy: 0.7769
Macro F1:          0.7919
Weighted F1:       0.8661
Macro Precision:   0.8188
Macro Recall:      0.7769
ROC-AUC (macro):   0.9427

Per-class metrics:
  forward    : F1=0.9160, Precision=0.8943, Recall=0.9387
  left       : F1=0.7034, Precision=0.8285, Recall=0.6111
  right      : F1=0.7565, Precision=0.7337, Recall=0.7807

Classification Report:
              precision    recall  f1-score   support

     forward       0.89      0.94      0.92      1469
        left       0.83      0.61      0.70       324
       right       0.73      0.78      0.76       187

    accuracy                           0.87      1980
   macro avg       0.82      0.78      0.79      1980
weighted avg       0.87      0.87      0.87      1980
```
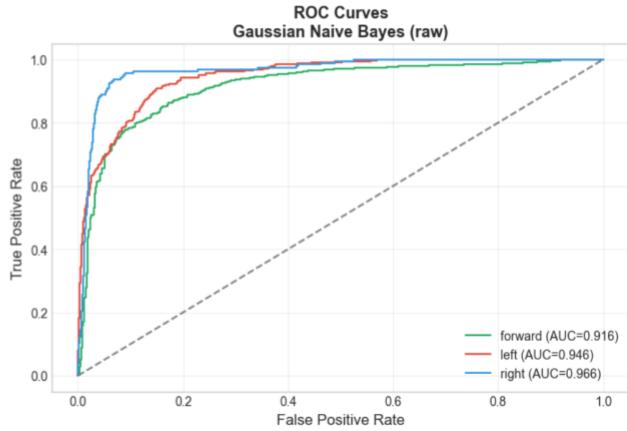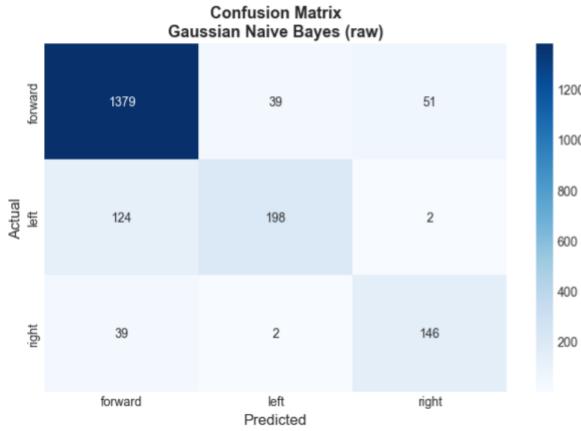

Confusion Matrix
Gaussian Naive Bayes (raw)

ROC Curves
Gaussian Naive Bayes (raw)

Engineered Features
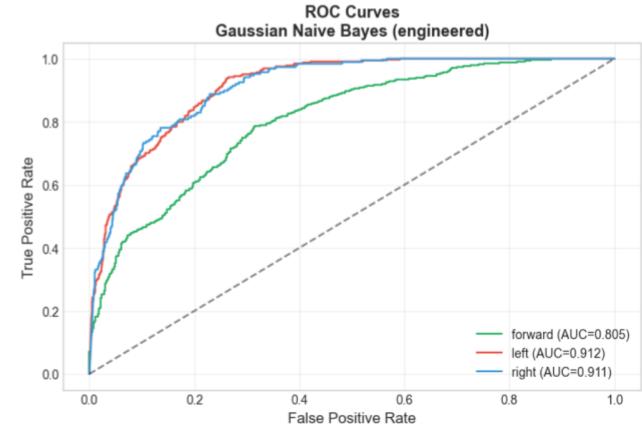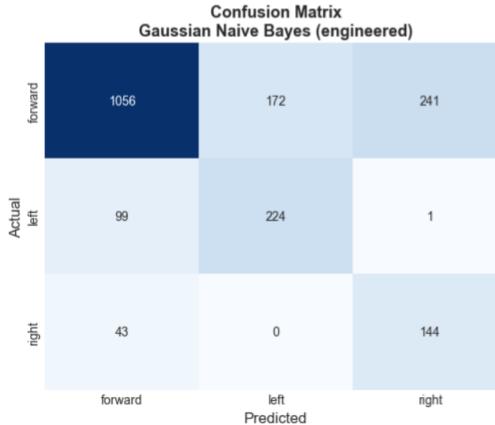
```
Training Gaussian Naive Bayes on ENGINEERED features...

============================================================
Model: Gaussian Naive Bayes | Features: engineered
============================================================
Accuracy:             0.7192
Balanced Accuracy: 0.7268
Macro F1:             0.6389
Weighted F1:          0.7368
Macro Precision:      0.6067
Macro Recall:         0.7268
ROC-AUC (macro):      0.8761

Per-class metrics:
  forward   : F1=0.7919, Precision=0.8815, Recall=0.7189
  left      : F1=0.6222, Precision=0.5657, Recall=0.6914
  right     : F1=0.5026, Precision=0.3731, Recall=0.7701

Classification Report:
             precision    recall  f1-score   support

    forward       0.88      0.72      0.79      1469
       left       0.57      0.69      0.62       324
      right       0.37      0.77      0.50       187

   accuracy                           0.72      1980
  macro avg       0.61      0.73      0.64      1980
weighted avg       0.78      0.72      0.74      1980
```


Confusion Matrix
Gaussian Naive Bayes (engineered)

ROC Curves
Gaussian Naive Bayes (engineered)

CART Trees

Raw Data
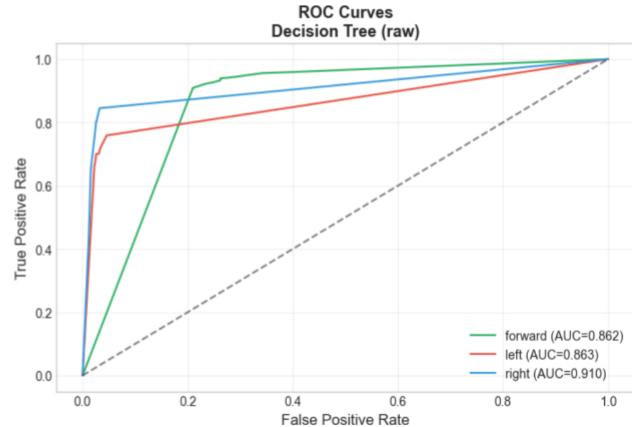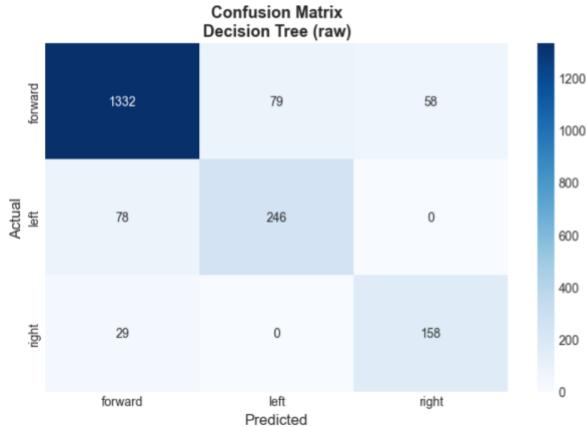
```
Evaluating best Decision Tree on RAW (PCA) features...

========================================================
Model: Decision Tree | Features: raw
========================================================
Accuracy:          0.8768
Balanced Accuracy: 0.8370
Macro F1:          0.8194
Weighted F1:       0.8778
Macro Precision:   0.8047
Macro Recall:      0.8370
ROC-AUC (macro):   0.8786

Per-class metrics:
  forward   : F1=0.9161, Precision=0.9256, Recall=0.9067
  left      : F1=0.7581, Precision=0.7569, Recall=0.7593
  right     : F1=0.7841, Precision=0.7315, Recall=0.8449

Classification Report:
             precision    recall  f1-score   support

     forward      0.93      0.91      0.92      1469
        left      0.76      0.76      0.76       324
       right      0.73      0.84      0.78       187

    accuracy                          0.88      1980
   macro avg      0.80      0.84      0.82      1980
weighted avg      0.88      0.88      0.88      1980
```


Confusion Matrix Decision Tree (raw)

ROC Curves Decision Tree (raw)

Engineered Features

```
Evaluating best Decision Tree on ENGINEERED features...

========================================================
Model: Decision Tree | Features: engineered
========================================================
Accuracy:          0.8384
Balanced Accuracy: 0.7697
Macro F1:          0.7591
Weighted F1:       0.8399
Macro Precision:   0.7494
Macro Recall:      0.7697
ROC-AUC (macro):   0.8256

Per-class metrics:
  forward   : F1=0.8899, Precision=0.8998, Recall=0.8802
  left      : F1=0.7034, Precision=0.6801, Recall=0.7284
  right     : F1=0.6841, Precision=0.6684, Recall=0.7005

Classification Report:
              precision    recall  f1-score   support

     forward       0.90      0.88      0.89      1469
        left       0.68      0.73      0.70       324
       right       0.67      0.70      0.68       187

    accuracy                           0.84      1980
   macro avg       0.75      0.77      0.76      1980
weighted avg       0.84      0.84      0.84      1980
```
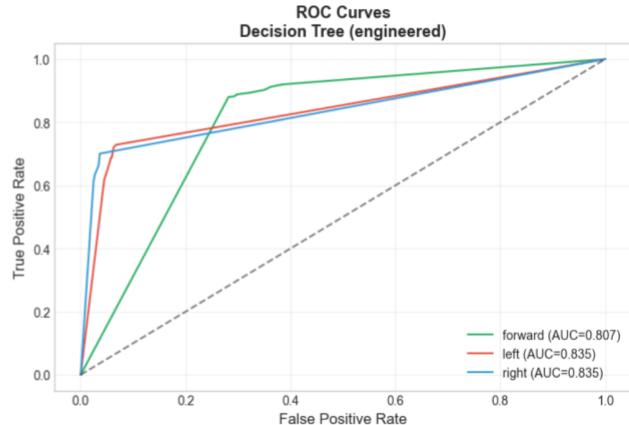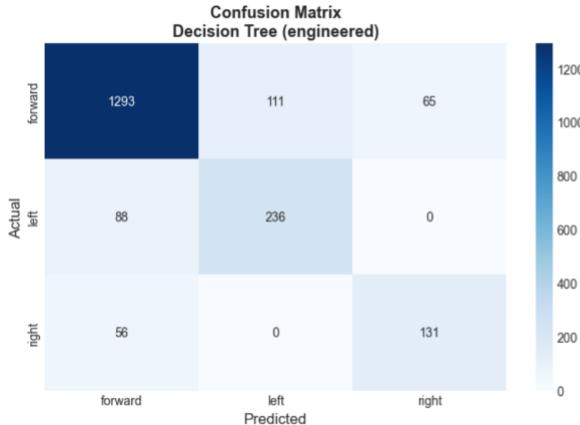

Confusion Matrix
Decision Tree (engineered)

ROC Curves
Decision Tree (engineered)

Bagging

Raw Data

```
Training Bagging Classifier on RAW (PCA) features...
OOB Score: 0.9186

=======================================================
Model: Bagging (Decision Trees) | Features: raw
=======================================================
Accuracy:          0.9172
Balanced Accuracy: 0.8513
Macro F1:          0.8703
Weighted F1:       0.9155
Macro Precision:   0.8932
Macro Recall:      0.8513
ROC-AUC (macro):   0.9758

Per-class metrics:
  forward   : F1=0.9452, Precision=0.9284, Recall=0.9626
  left      : F1=0.8233, Precision=0.8949, Recall=0.7623
  right     : F1=0.8424, Precision=0.8564, Recall=0.8289

Classification Report:
              precision    recall  f1-score   support

     forward       0.93      0.96      0.95      1469
        left       0.89      0.76      0.82       324
       right       0.86      0.83      0.84       187

    accuracy                           0.92      1980
   macro avg       0.89      0.85      0.87      1980
weighted avg       0.92      0.92      0.92      1980
```
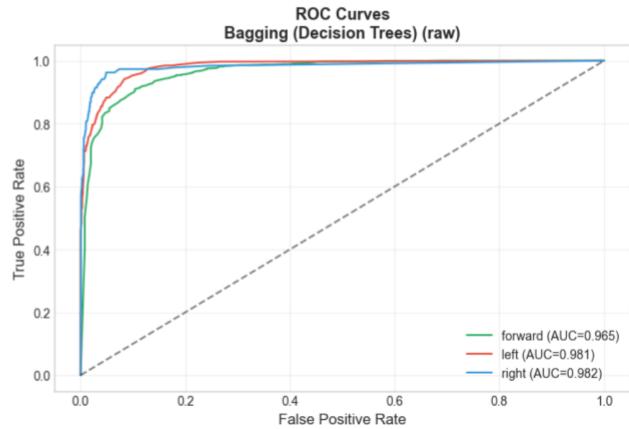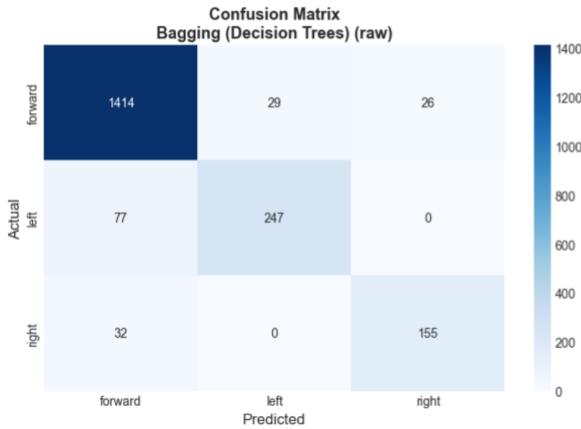

Confusion Matrix
Bagging (Decision Trees) (raw)

ROC Curves
Bagging (Decision Trees) (raw)
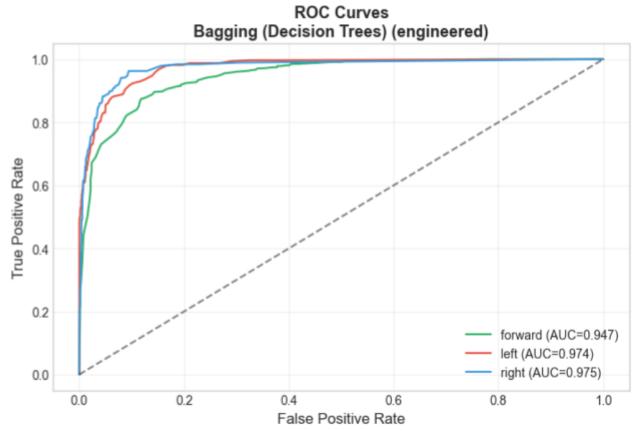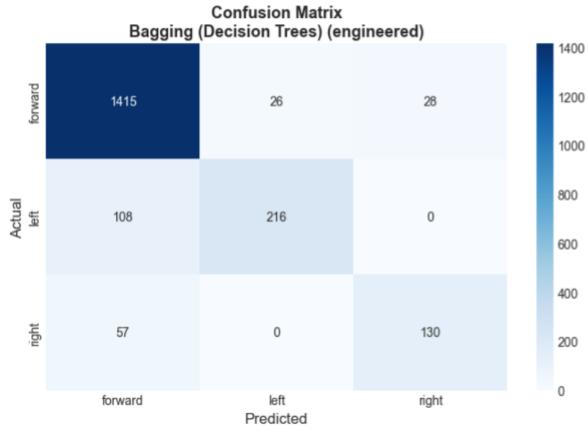
Engineered Features

```
Training Bagging Classifier on ENGINEERED features...
OOB Score: 0.8908

======================================================
Model: Bagging (Decision Trees) | Features: engineered
======================================================
Accuracy:           0.8894
Balanced Accuracy:  0.7750
Macro F1:           0.8150
Weighted F1:        0.8847
Macro Precision:    0.8703
Macro Recall:       0.7750
ROC-AUC (macro):    0.9649

Per-class metrics:
  forward  : F1=0.9282, Precision=0.8956, Recall=0.9632
  left     : F1=0.7633, Precision=0.8926, Recall=0.6667
  right    : F1=0.7536, Precision=0.8228, Recall=0.6952

Classification Report:
              precision    recall  f1-score   support

     forward       0.90      0.96      0.93      1469
        left       0.89      0.67      0.76       324
       right       0.82      0.70      0.75       187

    accuracy                           0.89      1980
   macro avg       0.87      0.78      0.82      1980
weighted avg       0.89      0.89      0.88      1980
```


Confusion Matrix
Bagging (Decision Trees) (engineered)

ROC Curves
Bagging (Decision Trees) (engineered)

Random Forest

Raw Data

```
Evaluating best Random Forest on RAW (PCA) features...
OOB Score: 0.9259

=======================================================
Model: Random Forest | Features: raw
=======================================================
Accuracy:           0.9232
Balanced Accuracy:  0.8554
Macro F1:           0.8796
Weighted F1:        0.9211
Macro Precision:    0.9111
Macro Recall:       0.8554
ROC-AUC (macro):    0.9809

Per-class metrics:
  forward    : F1=0.9495, Precision=0.9279, Recall=0.9721
  left       : F1=0.8253, Precision=0.9269, Recall=0.7438
  right      : F1=0.8641, Precision=0.8785, Recall=0.8503

Classification Report:
              precision    recall  f1-score   support

     forward       0.93      0.97      0.95      1469
        left       0.93      0.74      0.83       324
       right       0.88      0.85      0.86       187

    accuracy                           0.92      1980
   macro avg       0.91      0.86      0.88      1980
weighted avg       0.92      0.92      0.92      1980
```
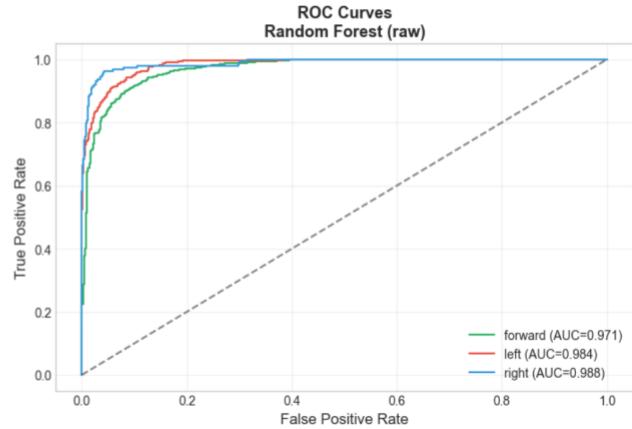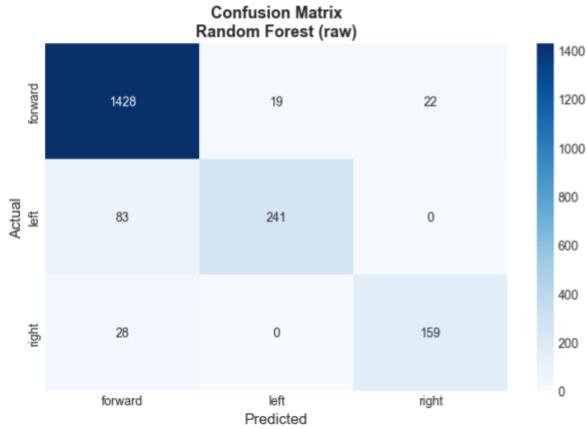

Confusion Matrix
Random Forest (raw)

ROC Curves
Random Forest (raw)

Engineered Features
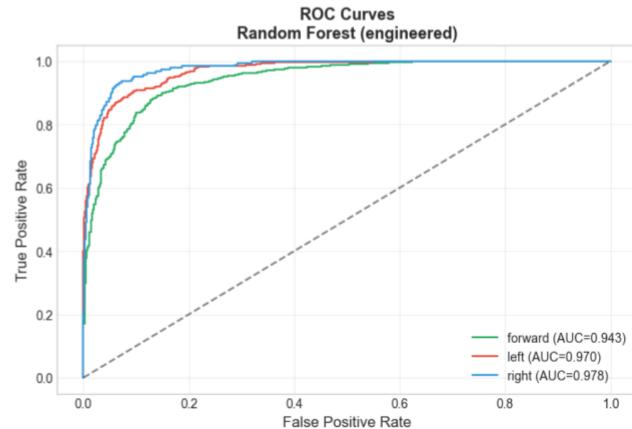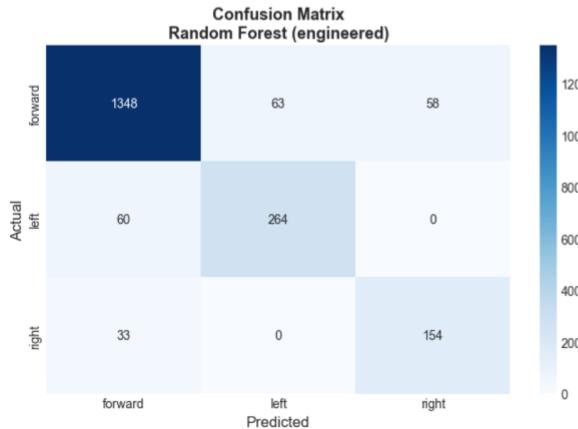
```
Evaluating best Random Forest on ENGINEERED features...
OOB Score: 0.8888

=======================================================
Model: Random Forest | Features: engineered
=======================================================
Accuracy:           0.8919
Balanced Accuracy:  0.8520
Macro F1:           0.8365
Weighted F1:        0.8930
Macro Precision:    0.8231
Macro Recall:       0.8520
ROC-AUC (macro):    0.9637

Per-class metrics:
  forward    : F1=0.9265, Precision=0.9355, Recall=0.9176
  left       : F1=0.8111, Precision=0.8073, Recall=0.8148
  right      : F1=0.7719, Precision=0.7264, Recall=0.8235

Classification Report:
              precision    recall  f1-score   support

     forward       0.94      0.92      0.93      1469
        left       0.81      0.81      0.81       324
       right       0.73      0.82      0.77       187

    accuracy                           0.89      1980
   macro avg       0.82      0.85      0.84      1980
weighted avg       0.89      0.89      0.89      1980
```


Confusion Matrix
Random Forest (engineered)

ROC Curves
Random Forest (engineered)

Boosting Methods

Raw Data
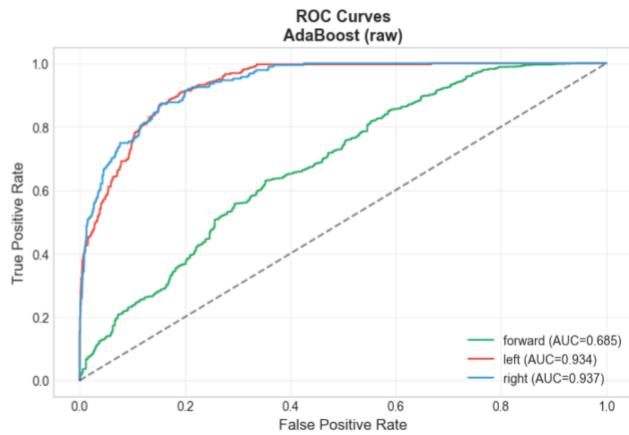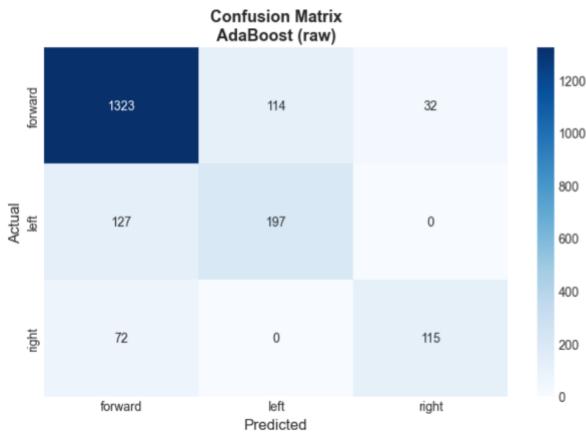
```
Training AdaBoost on RAW (PCA) features...

========================================================
Model: AdaBoost | Features: raw
========================================================
Accuracy:          0.8258
Balanced Accuracy: 0.7079
Macro F1:          0.7312
Weighted F1:       0.8229
Macro Precision:   0.7617
Macro Recall:      0.7079
ROC-AUC (macro):   0.8520

Per-class metrics:
  forward   : F1=0.8847, Precision=0.8693, Recall=0.9006
  left      : F1=0.6205, Precision=0.6334, Recall=0.6080
  right     : F1=0.6886, Precision=0.7823, Recall=0.6150

Classification Report:
              precision    recall  f1-score   support

     forward       0.87      0.90      0.88      1469
        left       0.63      0.61      0.62       324
       right       0.78      0.61      0.69       187

    accuracy                           0.83      1980
   macro avg       0.76      0.71      0.73      1980
weighted avg       0.82      0.83      0.82      1980
```


Confusion Matrix
AdaBoost (raw)

ROC Curves
AdaBoost (raw)

Engineered Features

```
Training Gradient Boosting on ENGINEERED features...

========================================================
Model: Gradient Boosting | Features: engineered
========================================================
Accuracy:           0.8960
Balanced Accuracy: 0.7914
Macro F1:           0.8261
Weighted F1:        0.8922
Macro Precision:    0.8735
Macro Recall:       0.7914
ROC-AUC (macro):    0.9653

Per-class metrics:
  forward  : F1=0.9325, Precision=0.9031, Recall=0.9639
  left     : F1=0.7852, Precision=0.9139, Recall=0.6883
  right    : F1=0.7606, Precision=0.8036, Recall=0.7219

Classification Report:
              precision    recall  f1-score   support

     forward       0.90      0.96      0.93      1469
        left       0.91      0.69      0.79       324
       right       0.80      0.72      0.76       187

    accuracy                           0.90      1980
   macro avg       0.87      0.79      0.83      1980
weighted avg       0.90      0.90      0.89      1980
```
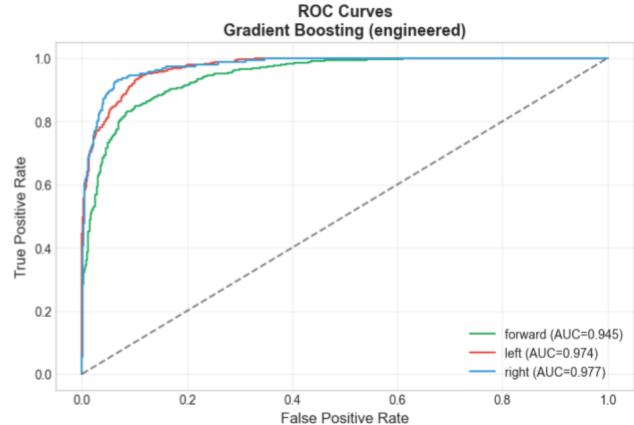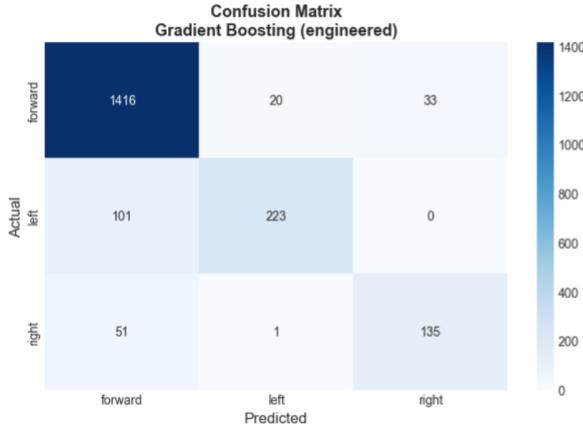

Confusion Matrix
Gradient Boosting (engineered)

ROC Curves
Gradient Boosting (engineered)

XG Boost

Raw Data
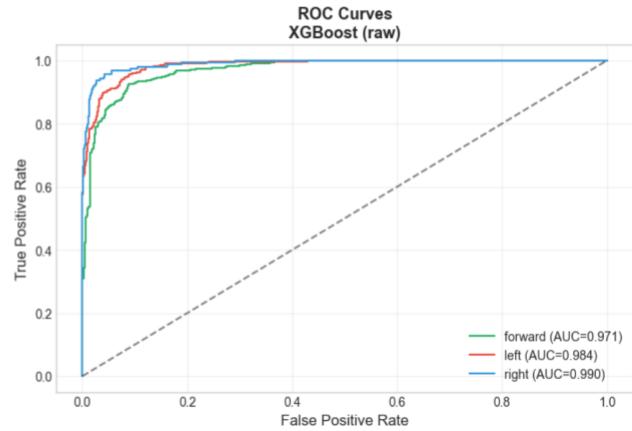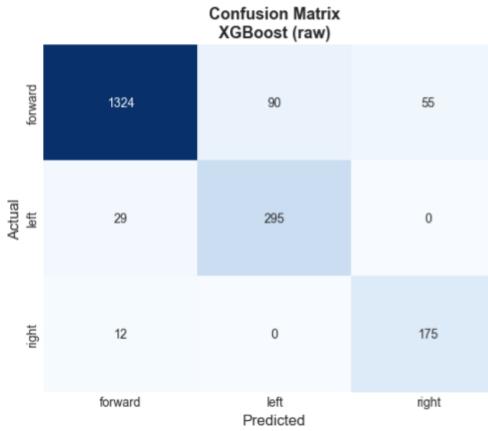
```
Evaluating best XGBoost on RAW (PCA) features...

========================================================
Model: XGBoost | Features: raw
========================================================
Accuracy:          0.9061
Balanced Accuracy: 0.9159
Macro F1:          0.8686
Weighted F1:       0.9087
Macro Precision:   0.8324
Macro Recall:      0.9159
ROC-AUC (macro):   0.9818

Per-class metrics:
  forward    : F1=0.9344, Precision=0.9700, Recall=0.9013
  left       : F1=0.8322, Precision=0.7662, Recall=0.9105
  right      : F1=0.8393, Precision=0.7609, Recall=0.9358

Classification Report:
              precision    recall  f1-score   support

     forward       0.97      0.90      0.93      1469
        left       0.77      0.91      0.83       324
       right       0.76      0.94      0.84       187

    accuracy                           0.91      1980
   macro avg       0.83      0.92      0.87      1980
weighted avg       0.92      0.91      0.91      1980
```


Confusion Matrix
XGBoost (raw)

ROC Curves
XGBoost (raw)

Engineered Features

```
Evaluating best XGBoost on ENGINEERED features...

========================================================
Model: XGBoost | Features: engineered
========================================================
Accuracy:          0.8727
Balanced Accuracy: 0.8843
Macro F1:          0.8239
Weighted F1:       0.8773
Macro Precision:   0.7821
Macro Recall:      0.8843
ROC-AUC (macro):   0.9680

Per-class metrics:
  forward    : F1=0.9099, Precision=0.9579, Recall=0.8666
  left       : F1=0.7901, Precision=0.7150, Recall=0.8827
  right      : F1=0.7717, Precision=0.6733, Recall=0.9037

Classification Report:
              precision    recall  f1-score   support

     forward       0.96      0.87      0.91      1469
        left       0.71      0.88      0.79       324
       right       0.67      0.90      0.77       187

    accuracy                           0.87      1980
   macro avg       0.78      0.88      0.82      1980
weighted avg       0.89      0.87      0.88      1980
```
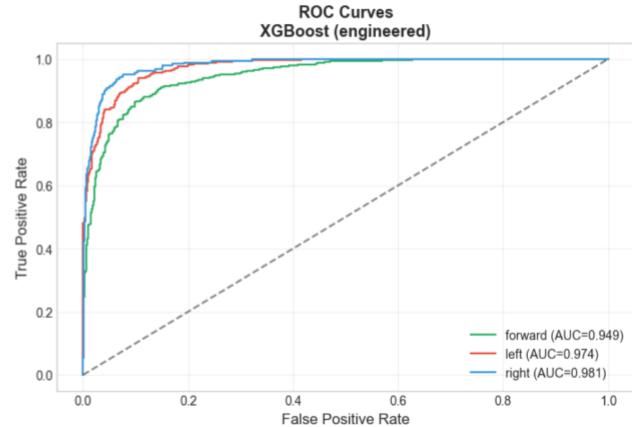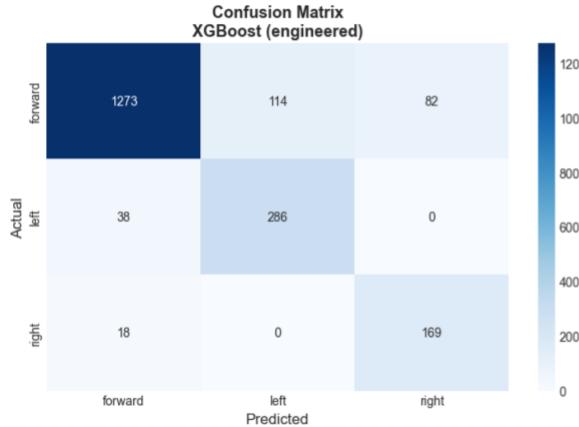

Confusion Matrix
XGBoost (engineered)

ROC Curves
XGBoost (engineered)

Support Vector Machines

Raw Data

```
Training Linear SVM on RAW (PCA) features...
Fitting 5 folds for each of 4 candidates, totalling 20 fits

Best C: 0.01
Best CV F1 score: 0.8627


========================================================
Model: SVM (Linear) | Features: raw
========================================================
Accuracy:          0.9040
Balanced Accuracy: 0.9285
Macro F1:          0.8647
Weighted F1:       0.9075
Macro Precision:   0.8200
Macro Recall:      0.9285
ROC-AUC (macro):   0.9874

Per-class metrics:
    forward  : F1=0.9321, Precision=0.9812, Recall=0.8877
    left     : F1=0.8528, Precision=0.7692, Recall=0.9568
    right    : F1=0.8092, Precision=0.7097, Recall=0.9412

Classification Report:
              precision    recall  f1-score   support

     forward       0.98      0.89      0.93      1469
        left       0.77      0.96      0.85       324
       right       0.71      0.94      0.81       187

    accuracy                           0.90      1980
   macro avg       0.82      0.93      0.86      1980
weighted avg       0.92      0.90      0.91      1980
```
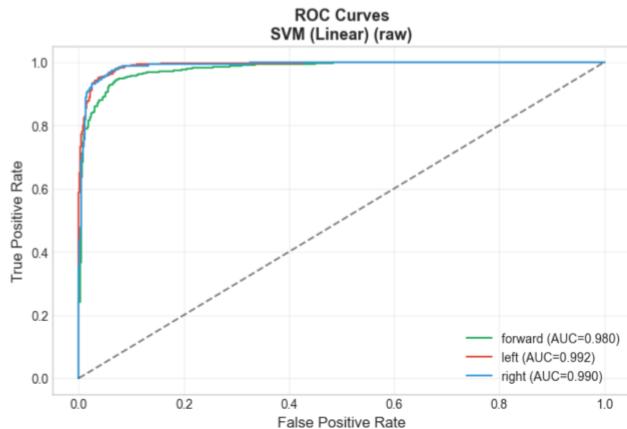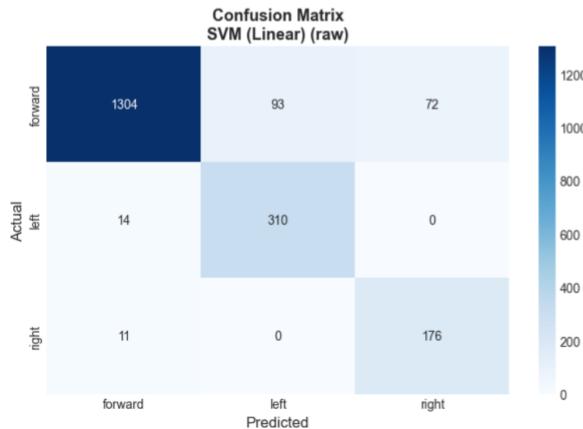

Confusion Matrix
SVM (Linear) (raw)

ROC Curves
SVM (Linear) (raw)

Engineered Features

```
Training Linear SVM on ENGINEERED features...
Fitting 5 folds for each of 4 candidates, totalling 20 fits

Best C: 10
Best CV F1 score: 0.7557

=======================================================
Model: SVM (Linear) | Features: engineered
=======================================================
Accuracy:           0.8086
Balanced Accuracy:  0.8664
Macro F1:           0.7554
Weighted F1:        0.8195
Macro Precision:    0.7038
Macro Recall:       0.8664
ROC-AUC (macro):    0.9482

Per-class metrics:
  forward   : F1=0.8574, Precision=0.9588, Recall=0.7754
  left      : F1=0.7268, Precision=0.6177, Recall=0.8827
  right     : F1=0.6822, Precision=0.5350, Recall=0.9412

Classification Report:
              precision    recall  f1-score   support

     forward       0.96      0.78      0.86      1469
        left       0.62      0.88      0.73       324
       right       0.53      0.94      0.68       187

    accuracy                           0.81      1980
   macro avg       0.70      0.87      0.76      1980
weighted avg       0.86      0.81      0.82      1980
```
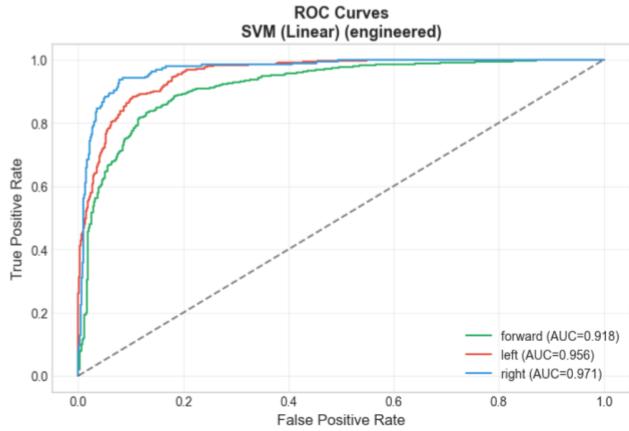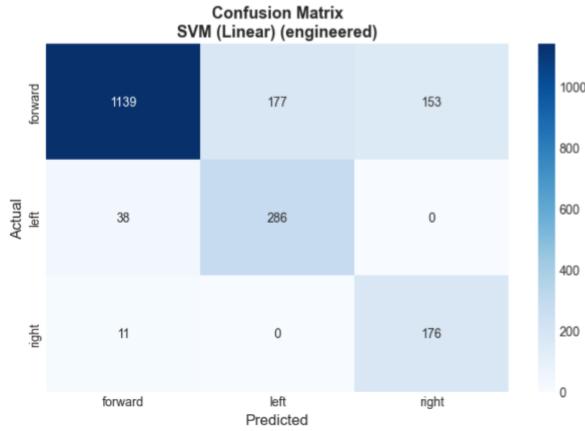

Confusion Matrix
SVM (Linear) (engineered)

ROC Curves
SVM (Linear) (engineered)

Model Summaries

	Model	Features	Accuracy	Balanced Accuracy	F1 (Macro)	F1 (Weighted)	Precision (Macro)	Recall (Macro)	ROC-AUC
0	Decision Tree	raw	0.876768	0.836973	0.819434	0.877774	0.804682	0.836973	0.878620
1	Decision Tree	engineered	0.838384	0.769707	0.759128	0.839935	0.749425	0.769707	0.825643
2	XGBoost	raw	0.906061	0.915872	0.868618	0.908667	0.832356	0.915872	0.981808
3	XGBoost	engineered	0.872727	0.884345	0.823893	0.877262	0.782057	0.884345	0.968032
4	kNN (k=5)	raw	0.925253	0.872437	0.885235	0.924492	0.899079	0.872437	0.965739
5	kNN (k=7)	engineered	0.872727	0.738162	0.789383	0.865492	0.866511	0.738162	0.940597
6	Logistic Regression (Standard)	raw	0.909596	0.925576	0.868652	0.912677	0.827419	0.925576	0.987109
7	Logistic Regression (Standard)	engineered	0.801515	0.853810	0.744852	0.813044	0.695001	0.853810	0.946896
8	Logistic Regression (L1/Lasso)	raw	0.910101	0.927358	0.869894	0.913159	0.828327	0.927358	0.987304
9	Logistic Regression (L1/Lasso)	engineered	0.802525	0.853462	0.745890	0.813854	0.696025	0.853462	0.946795
10	Logistic Regression (L2/Ridge)	raw	0.908586	0.924320	0.867686	0.911664	0.826578	0.924320	0.987376
11	Logistic Regression (L2/Ridge)	engineered	0.803030	0.856046	0.746940	0.814401	0.696793	0.856046	0.947006
12	Random Forest	raw	0.923232	0.855395	0.879647	0.921097	0.911084	0.855395	0.980903
13	Random Forest	engineered	0.891919	0.851992	0.836483	0.892982	0.823072	0.851992	0.963663
14	Bagging (Decision Trees)	raw	0.917172	0.851261	0.870304	0.915539	0.893237	0.851261	0.975784
15	Bagging (Decision Trees)	engineered	0.889394	0.775031	0.815016	0.884701	0.870305	0.775031	0.964913
16	Gaussian Naive Bayes	raw	0.870202	0.776865	0.791942	0.866122	0.818804	0.776865	0.942689
17	Gaussian Naive Bayes	engineered	0.719192	0.726756	0.638914	0.736814	0.606728	0.726756	0.876118
18	AdaBoost	raw	0.825758	0.707870	0.731250	0.822910	0.761668	0.707870	0.851950
19	AdaBoost	engineered	0.601515	0.721036	0.562007	0.632690	0.556883	0.721036	0.848059
20	Gradient Boosting	raw	0.918182	0.835019	0.870628	0.915177	0.917100	0.835019	0.976320
21	Gradient Boosting	engineered	0.895960	0.791373	0.826091	0.892159	0.873522	0.791373	0.965293
22	SVM (Linear)	raw	0.904040	0.928548	0.864703	0.907515	0.820032	0.928548	0.987379
23	SVM (Linear)	engineered	0.808586	0.866417	0.755446	0.819450	0.703806	0.866417	0.948223
24	SVM (RBF)	raw	0.917172	0.923462	0.880282	0.919335	0.846309	0.923462	0.987613
25	SVM (RBF)	engineered	0.871717	0.887051	0.823576	0.876471	0.780392	0.887051	0.967777

Model Comparison: F1 Macro

Model Comparison: Balanced Accuracy

Legend: Raw (PCA), Engineered

Model	Raw (PCA)	Engineered
Decision Tree		0.770
XGBoost	0.916	0.884
kNN (k=5)	0.872	
kNN (k=7)		0.738
Logistic Regression (Standard)	0.926	0.854
Logistic Regression (L1/Lasso)	0.927	0.853
Logistic Regression (L2/Ridge)	0.924	0.856
Random Forest	0.855	0.852
Bagging (Decision Trees)	0.851	0.775
Gaussian Naive Bayes	0.777	0.727
AdaBoost	0.708	0.721
Gradient Boosting	0.835	0.791
SVM (Linear)	0.929	0.866
SVM (RBF)	0.923	0.887