Step 1 : **Data Collection** – collected over around 40,000 images of which contain both drowsy and non-drowsy images from Kaggle.

Step 2 **: Data Preprocessing** – cleaned and prepared the data for feature extraction.

* Resized all images for 128 \* 128-pixel size.
* Converted all images to Grayscale as we don’t need the color data from images
* Normalized the pixel values to 0-1 range as we need to have same intensity levels across the data set.

Step 3 : **Feature Extraction** – Extracted 11 features from image data, utilized dlib’s landmark detection points to extract these features.

A blue dots with numbers

Description automatically generated

facial landmark detection was performed using **Dlib's 68-point facial landmark model**, which employs a pre-trained **Histogram of Oriented Gradients (HOG)** detector and **ensemble regression trees** for precise feature point localization on faces. These landmark points were used to extract meaningful features like Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), head tilt, and other facial metrics critical for drowsiness detection.

In Simple Terms:

* **HOG detector**: Finds the basic shapes and edges in the image by analyzing how brightness changes.
* **Ensemble regression trees**: A group of mini algorithms that work together to pinpoint facial features based on the patterns detected by HOG.

*Left EAR (Eye Aspect Ratio):*

Measures vertical vs horizontal eye distance for the left eye to detect closure/blinking.

Dlib Points: 36-41.

*Right EAR (Eye Aspect Ratio):*

Same as Left EAR but for the right eye.

Dlib Points: 42-47.

*Avg EAR (Average Eye Aspect Ratio):*

Average of Left and Right EAR values.

Helps generalize eye closure detection.

*MAR (Mouth Aspect Ratio):*

Measures vertical mouth opening vs horizontal width to detect yawning.

Dlib Points: 48-67.

*Lip Distance:*

Vertical distance between upper and lower lips to detect yawning.

Dlib Points: p13, p19.

*Mouth Width-Height Ratio:*

Ratio of mouth width to vertical lip distance, identifying mouth opening behavior.

Dlib Points: p12, p16, p13, p19.

*Head Tilt:*

Measures the tilt angle of the head using the line between eye corners.

Dlib Points: 36, 45.

*Head Nod:*

Measures the vertical angle between the chin and nose tip to detect head nodding.

Dlib Points: 30, 8.

*Eyebrow Distance:*

Average distance between eyebrow landmarks and corresponding eye landmarks.

Dlib Points: 17-21, 22-26 (eyebrows), 37, 44 (eyes).

*Eyebrow Asymmetry:*

Difference in average distances between the left and right eyebrows and eyes.

Indicates uneven facial behavior due to fatigue.

*Facial Droop Asymmetry:*

Vertical difference between left and right mouth corners, capturing facial droop.

Dlib Points: 48, 54.

Step 4 – **Feature Engineering and Dataset Preparation** -- Structured extracted features into a DataFrame using Pandas. Each row represents an image, and each column represents a feature. Each image is labeled with a binary value indicating the driver’s state (drowsy = 1, non-drowsy = 0).

Step 5 - **Model Development –** trained a Random Forest Model to classify drowsy vs non-drowsy drivers based on extracted features, the main goal is to get the most important features that contribute to the decision-making process.

* Divided the dataset into a training set (e.g., 80%) and a test set (e.g., 20%).
* Used Random Forest model for classification.
* Used cross-validation to tune hyperparameters and avoid overfitting.

Step 6 – **Model Evaluation** - evaluated the performance of trained models by calculating metrics like accuracy, precision and more as follows.

Random Forest Model Evaluation:

**Accuracy**: 0.920384718140529

**Confusion Matrix:**

**(Non-drowsy)0 1(drowsy)**

[[ 3231 338]

[ 258 3659]]

True Positives (TP): 3659

The model correctly predicted "drowsy" for 3659 samples.

True Negatives (TN): 3231

The model correctly predicted "non-drowsy" for 3231 samples.

False Positives (FP): 338

The model incorrectly predicted "drowsy" for 338 samples that were actually "non-drowsy."

False Negatives (FN): 258

The model incorrectly predicted "non-drowsy" for 258 samples that were actually "drowsy."

**Classification Report:**

Class precision recall f1-score support

0 0.93 0.91 0.92 3569

1 0.92 0.93 0.92 3917

accuracy 0.92 7486

macro avg 0.92 0.92 0.92 7486

weighted avg 0.92 0.92 0.92 7486

Key Metrics

1. Accuracy

The overall percentage of correct predictions.

Interpretation: The model correctly classified 92% of the samples.

2. Precision

What it tells you: Out of all the samples predicted as "drowsy" (class 1), how many were actually "drowsy."

Interpretation:

When the model predicts "drowsy," it is correct 92% of the time.

When it predicts "non-drowsy," it is correct 93% of the time.

3. Recall

What it tells you: Out of all the actual "drowsy" samples, how many did the model correctly identify.

Interpretation:

The model correctly identified 93% of the "drowsy" samples.

It correctly identified 91% of the "non-drowsy" samples.

4. F1-Score:

What it tells you: The harmonic mean of precision and recall, balancing the trade-off between false positives and false negatives.

Interpretation: The F1-Score is a good balance of precision and recall, indicating that the model performs well at distinguishing between drowsy and non-drowsy states.

5. Macro Average:

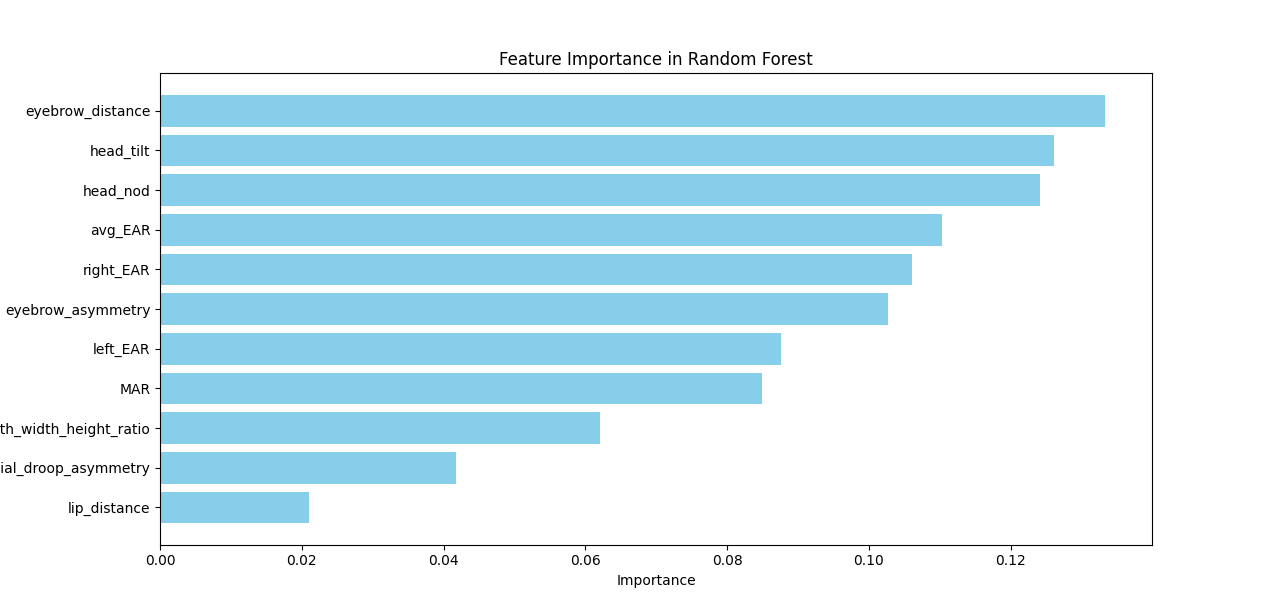
The average of precision, recall, and F1-scores across both classes (0 and 1), treating both classes equally.

Interpretation: It shows the model’s performance across classes without considering class imbalance.

6. Weighted Average:

The weighted average of precision, recall, and F1-scores, considering the number of samples in each class.

Interpretation: Since class 1 ("drowsy") has slightly more samples, the weighted average reflects the model's performance more closely for this class.



While building the trees, the forest keeps track of how much each feature contributes to reducing impurity.

After training, it calculates Gini Importance for each feature:

For each tree, it adds up the impurity reduction caused by each feature.

It averages these values across all trees to get the overall importance of each feature.

A graph of a graph

Description automatically generated with medium confidence

The SHAP plot shows how these numeric values influence predictions:

Red for higher values often pushes the prediction toward drowsy or increases the likelihood of being drowsy.

Blue for lower values often reduces the likelihood of being drowsy or pushes the prediction toward non-drowsy.