MLPR END TERM PRESENTATION

AADI JAIN
VAIBHAV DABAS
MOHIL AHUJA

Real Time Drowsiness Detection in Cases of *Occlusion*

Built for light-weight, low compute power devices.

PROBLEM

Now Playing Top News of the Day

NEWS / CITY NEWS / VISAKHAPATNAM NEWS / Driver Fatigue Related Accidents On The Rise

TRENDING Shimla Mosque Row Mumbai Marine Drive Hyderabad Junior Doctor Assault Kolkata

Driver fatigue related accidents on the rise

Nalla.Babu / Aug 11, 2024, 05:03 IST SHARE

Delhi-Meerut Expressway road accident: Why it is a lesson for highway drivers

By: HT Auto Desk | Updated on: 02 Apr 2024, 16:40 PM Share via:

Follow us on:

Fatigue, highway hypnosis and other factors may often lead drivers to dose off behind the wheels or take eyes off the road, putting passengers and oth ... Read more

HOME / NEWS / INDIA / KERALA

Sleep-deprived drivers responsible for 40% of road accidents, say transport officials

Fatigued drivers are prone to dozing off at the wheels at night, according to them

Updated - November 28, 2021 11:26 am IST - KOCHI:

G. KRISHNAKUMAR

17.6%

Of all fatal car crashes

between 2017 and 2021 involved a **drowsy driver (AAA Foundation)**

$109 billion

annually

is the cost of **fatigue-related** collisions. (**NHTSA**)

https://aaafoundation.org/drowsy-driving-in-fatal-crashes-united-states-2017-2021/

https://www.nhtsa.gov/risky-driving/drowsy-driving

PROBLEM STATEMENT

Problem Statement:

We want to detect driver drowsiness even in *occlusion conditions*.

Real Time Drowsines
Detection in Cases o

What attempts have been made?

and how we’re aiming for something new

AS OF 2023...

MAJEED ET AL. (2023)

Majeed F, Shafique U, Safran M, Alfarhood S, Ashraf I. Detection of Drowsiness among Drivers Using Novel Deep Convolutional Neural Network Model. Sensors. 2023; 23(21):8741. https://doi.org/10.3390/s23218741

THE FIELD SO FAR

Ref.	Base Parameters	Model	Accuracy (Test)	Dataset (s)
[27]	Facial Regions	Fusion System (Analysis of Mixed Datasets)	93.38%, 96.68%	YawDD, DEAP, MiraclHB
[28]	Facial Regions	SVM+LSTM	89%	UTA-RLDD
[19]	Eyes and Mouth	Multi-physical Feature Fusion Detection Method Detection Method based on Deep Learning SSD+VGG16	95.7%, 91.4% (Custom), 91.88%	(Public) Homemade Dataset, NTHU-DDD
[38]	Facial Regions	Weibull-based MobileNetV2 Weibull-based ResNext101 MTCNN+Weibull Pooling+ResNext101	93.8%, 90.5%, 84.21%	Custom Dataset (Total 50, 30 male, 20 female) NTHU-DDD
[29]	Facial Regions	2-stream spatial-temporal graph convolutional network (2s-STGCN)	93.4%, 92.7%	YawDD, NTHU-DDD
[35]	Facial Regions	3D Deep CNN CNN (LeNet)	96.80%	UTA-RLDD Custom Dataset (10 Subjects)
[30]	Facial Regions	Linear Support Vector Machine (SVM) as classifier + Dlib	92.5%	YawDD
[40]	Facial Regions	SVM + Dlib facial feature predictor	94.55%	IMM face Dataset + Other Mixed Samples
[31]	Facial Regions	TFBI LSTM, CNN-LSTM	79.9% Temporal, 97.5% Spatial	UTA-RLDD
[34]	Mouth and Eyes	MTCNN+DLIB+LSTM NN	88%, 90%	YawDD Self-Built Dataset
[32]	Eyes and Mouth	SVM and Adaboost + Multitask ConNN	98.81%	YawDD and NthuDDD
[36]	Facial Regions	YOLOv3-tiny CNN + Face Feature, Triangle (FFT) + Face Feature Vector (FFV)	94.32%	YawDD
[37]	Facial Regions	Conv2D-raw + SMOTE	64%	Real-Time Generated Dataset
[33]	Facial Regions	3DcGAN+TLABiLSTM+Refinement, 3DcGAN+TLABiLSTM, 3DcGAN	91.20%, 87.1%, 82.8%	NthuDDD
[39]	Eyes	HM-LSTM network, LSTM network	65.2%, 61.4%	Custom Dataset

TAMANANI ET AL. (2023)

R. Tamanani, R. Muresan and A. Al-Dweik, "Estimation of Driver Vigilance Status Using Real-Time Facial Expression and Deep Learning," in IEEE Sensors Letters, vol. 5, no. 5, pp. 1-4, May 2021, Art no. 6000904, doi: 10.1109/LSENS.2021.3070419.

THEY DID BUILD SOMETHING LIGHT,
BUT THERE'S SCOPE FOR IMPROVEMENT.

91% ACCURACY

[35]	Facial Regions	3D Deep CNN CNN (LeNet)	96.80%	UTA-RLDD Custom Dataset (10 Subjects)

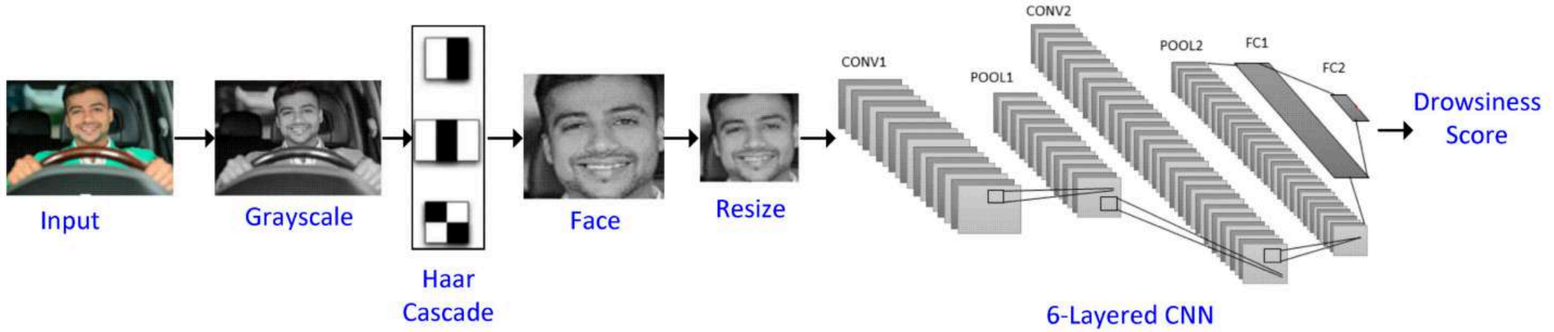
IEEE Sensors Letters VOL. 5, NO. 5, MAY 2021 6000904 IEEE Sensors Council

Sensor applications

Estimation of Driver Vigilance Status Using Real-Time Facial Expression and Deep Learning

Reza Tamanani[1]*, Radu Muresan[1]**, and Arafat Al-Dweik[1,2]**

[1]School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada



on UTA-RLDD showed that the model achieved high values of average accuracy, precision, recall, and F1-score, which are 0.918, 0.928, 0.920, and 0.920, respectively. Moreover, comparison results showed that the proposed system outperforms other methods with the same

QU ET AL. (2023)

Gao, Z. (2023). Multi-Attention Fusion Drowsy Driving Detection Model. arXiv preprint arXiv:2312.17052. https://arxiv.org/abs/2312.17052

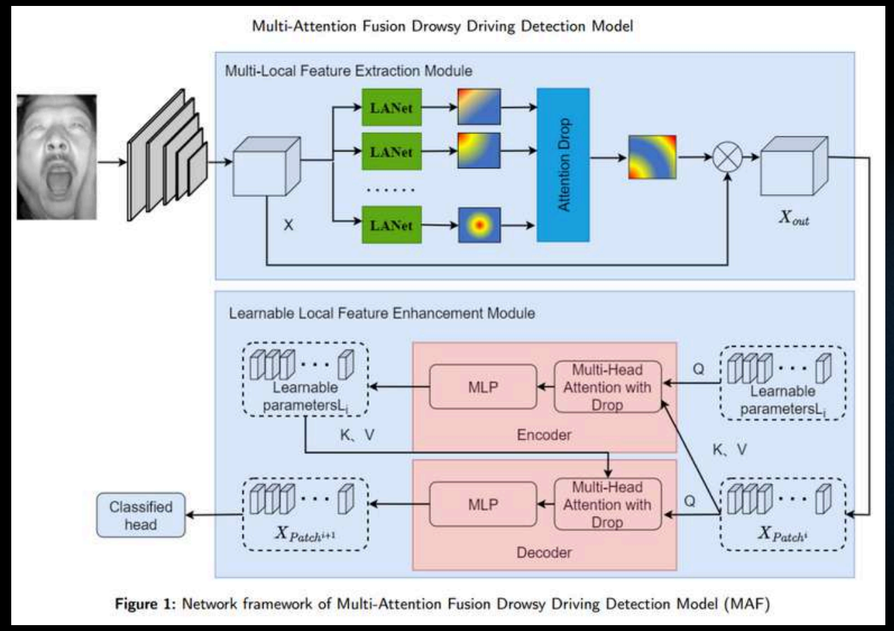


Figure 1: Network framework of Multi-Attention Fusion Drowsy Driving Detection Model (MAF)

Multi-Attention Fusion Drowsy Driving Detection Model

Shulei QU[a], Zhenguo Gao[a,*], Xiaoxiao Wu[a] and Yuanyuan Qiu[a]

[a]*School of Computer Science and Technology, Huaqiao University, China*

ARTICLE INFO

Keywords:
Drowsiness detection
Fatigue detection
CNN
Attention Fusion

ABSTRACT

Drowsy driving represents a major contributor to traffic accidents, and the implementation of driver drowsy driving detection systems has been proven to significantly reduce the occurrence of such accidents. Despite the development of numerous drowsy driving detection algorithms, many of them impose specific prerequisites such as the availability of complete facial images, optimal lighting conditions, and the use of RGB images. In our study, we introduce a novel approach called the Multi-Attention Fusion Drowsy Driving Detection Model (MAF). MAF is aimed at significantly enhancing

IMPRESSIVE ACCURACY, BUT HIGH COST, HIGH PARAMETER COUNTS

ATTENTION BASED ARCHITECTURE, WITH RESNET50 BACKBONE

Parameter Count

ResNet-50 is trained on over a million images from the ImageNet dataset, which consists of more than 14 million images across 1000 classes. The total number of parameters in ResNet-50 is approximately 25.6 million. This relatively low parameter count, combined with its depth, allows ResNet-50 to achieve high accuracy in various computer vision tasks while maintaining

The Dataset

UTA-Real Life Drowsiness Detection (RLDD)

Real Time Drowsines
Detection in Cases o
Occlusion

DATASET AND FEATURES PREPROCESSING

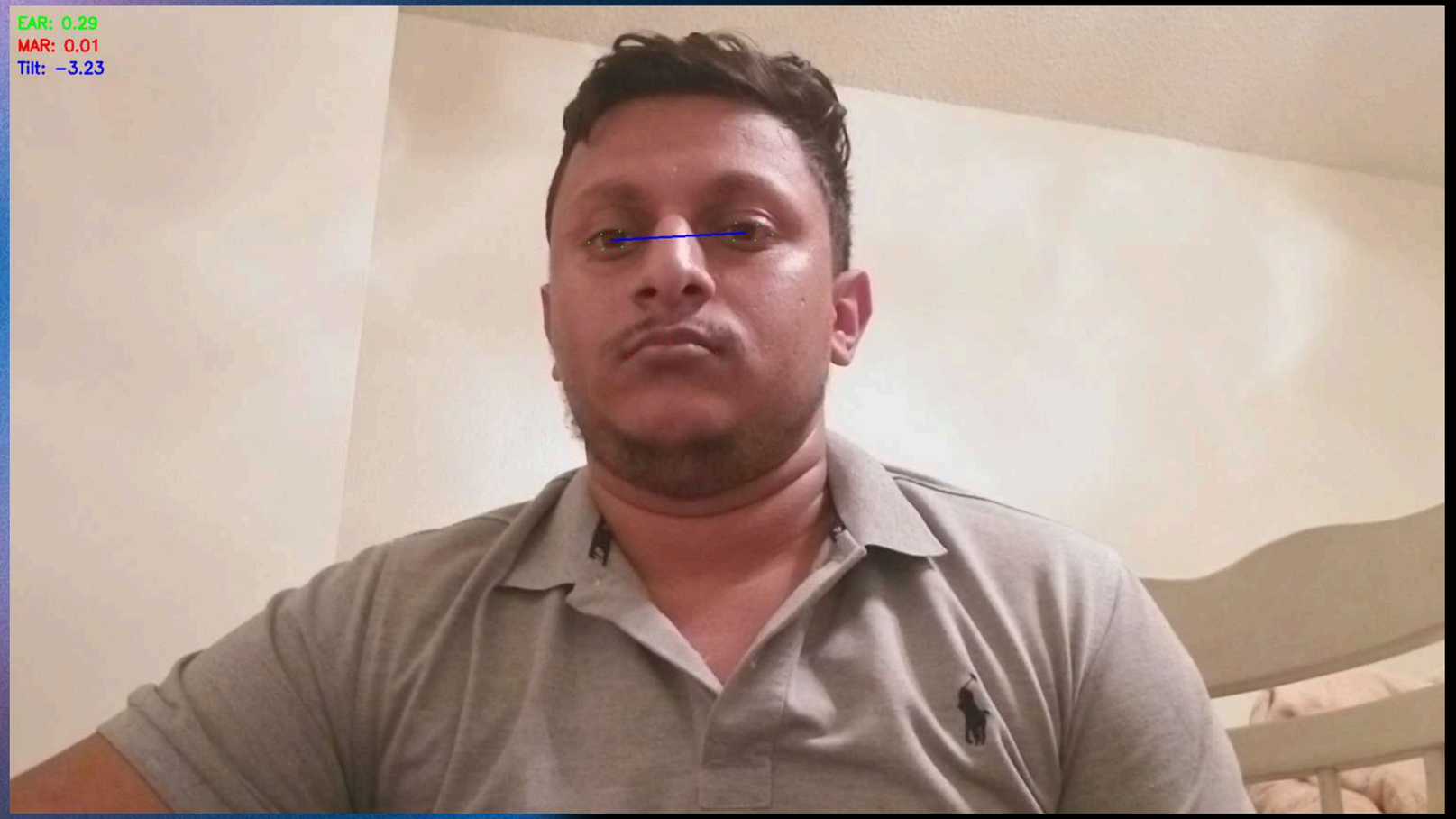
(UTA-RLDD) was created for the task of multi-stage drowsiness detection, targeting not only extreme and easily visible cases, but also subtle cases when subtle micro-expressions are the discriminative factors.

Each participant provided one video per class: alert, low vigilance, and drowsy.

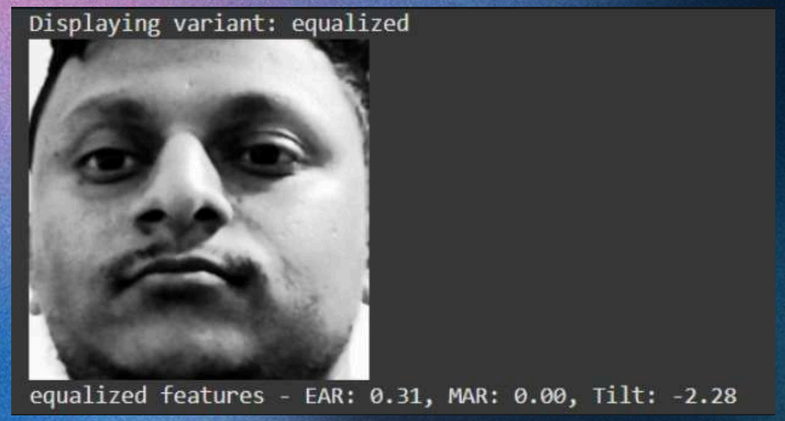
- **Dataset:** UTA-RLDD for real-life drowsiness detection
- **Nature:** 180 videos, 60 participants, diverse demographics.
- **Why:** Includes partial occlusion cases, realistic scenarios, early detection focus
- **Data Collection:** Self-recorded videos in various environments.
- **Ethical Concerns:** Anonymity ensured, voluntary participation + Extra credit
- **Features & Datapoints**: Facial expressions, eye and mouth movements, 10 minutes at a time
- **Occlusion:** Glasses or considerable facial hair in 50% of the dataset.
- **Total Features & Datapoints**: 180 (10 minute long) videos with multiple facial features

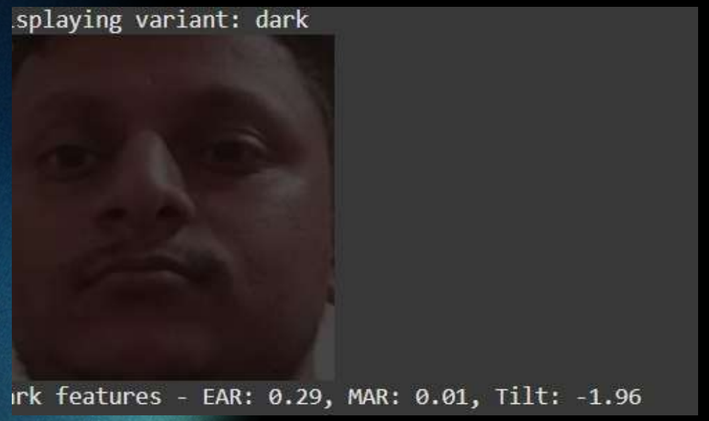
R. Ghoddoosian, M. Galib and V. Athitsos, "A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019, pp. 178-187, doi: 10.1109/CVPRW.2019.00027.

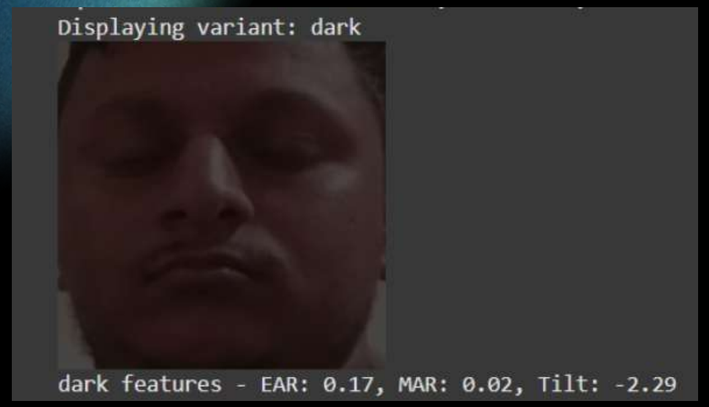
PREPROCESSING: EAR, MAR, TILT VIA MEDIAPIPE

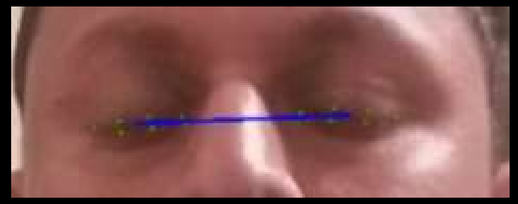


PREPROCESSING: CROPPING, SHADOWS, SYNTHETIC OCCLUSIONS

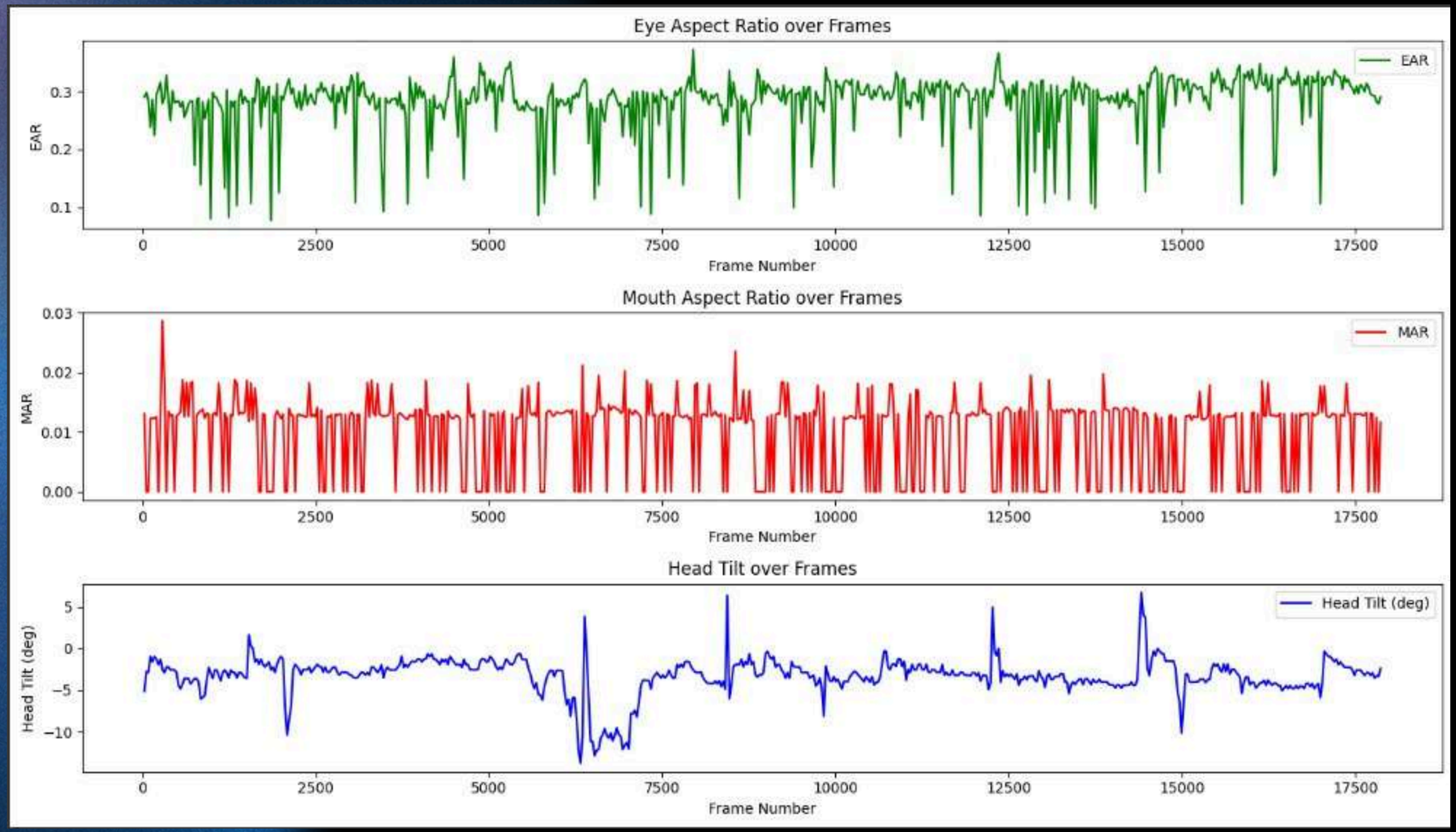




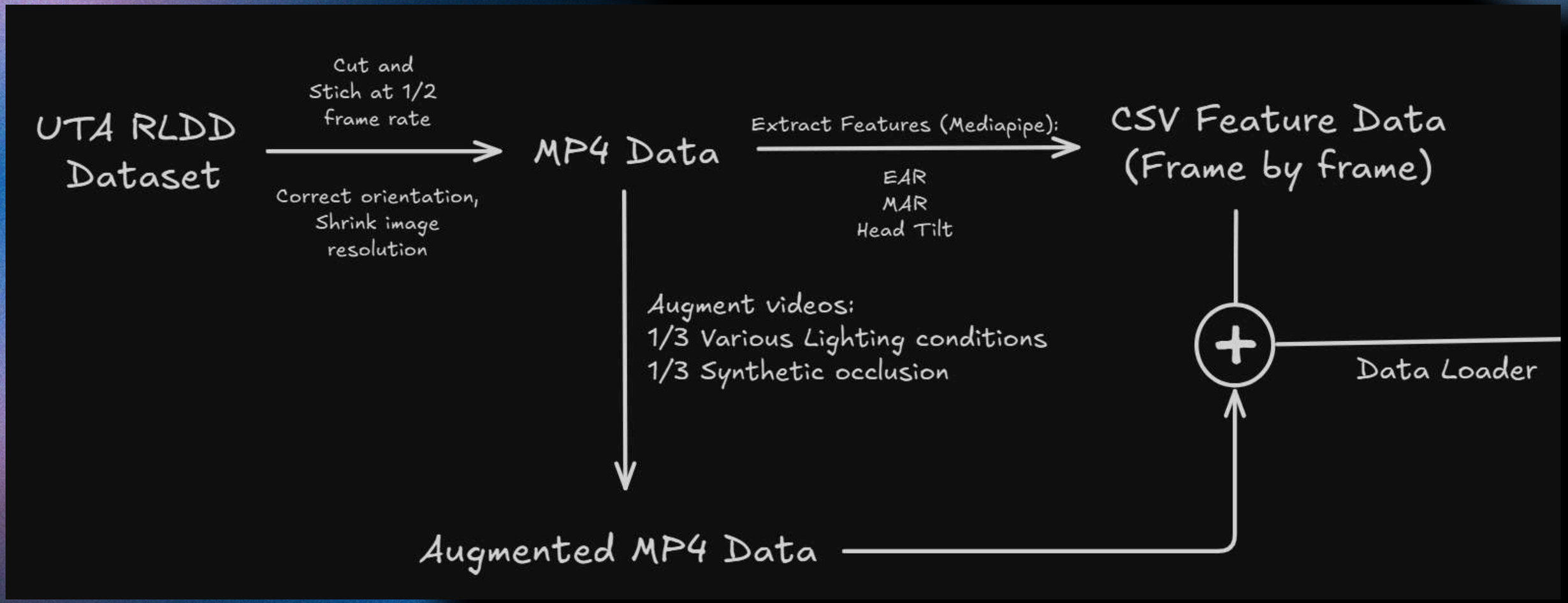




FULL VIDEO GRAPH



AN OVERVIEW OF THE PREPROCESSING



WHAT WE USED

ML Methodology

WE TRIED:

1. DENSE NEURAL NETWORK (ON EXTRACTED DATA)
2. STANDARD CNN WITH FULLY PREPROCESSED DATA
3. LSTM (ON EXTRACTED DATA)
4. CNN + LSTM WITH FULLY PREPROCESSED DATA

WHAT WE USED

ML Methodology

TRIALS & FAILURES

- DENSE NEURAL NETWORK (ON EXTRACTED DATA)

  ACCURACY : 61%

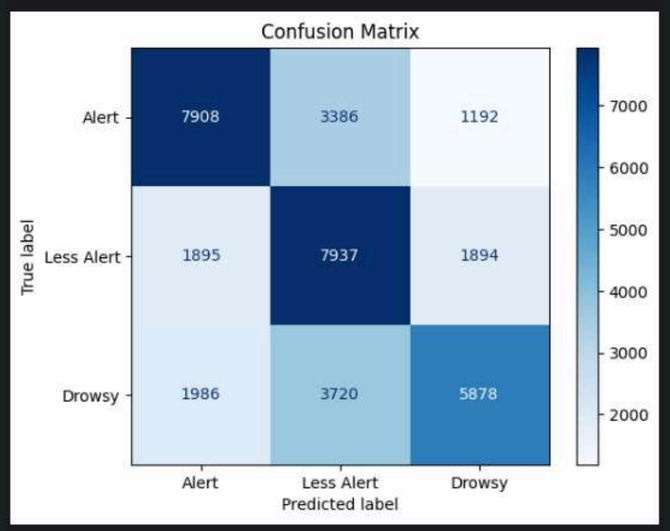
  MODEL NOT POWERFUL ENOUGH

- STANDARD CNN WITH FULLY PREPROCESSED DATA

  ACCURACY : 68%

  NO MANUAL FEATURES, NO TEMPORAL CAPTURE

```
Test Accuracy: 0.61
1119/1119 ━━━━━━━━━━━━━━━━━━━━ 1s 1ms/step
```



	precision	recall	f1-score	support
Alert	0.67	0.63	0.65	12486
Less Alert	0.53	0.68	0.59	11726
Drowsy	0.66	0.51	0.57	11584

WHAT WE USED

- **CNN + LSTM + MEDIAPIPE FEATURES**

  ACCURACY : 83%

  LSTM UNABLE TO TAKE LONG CONTEXT LENGTHS, LEADING TO INSUFFICIENT RUNTIME ON KAGGLE

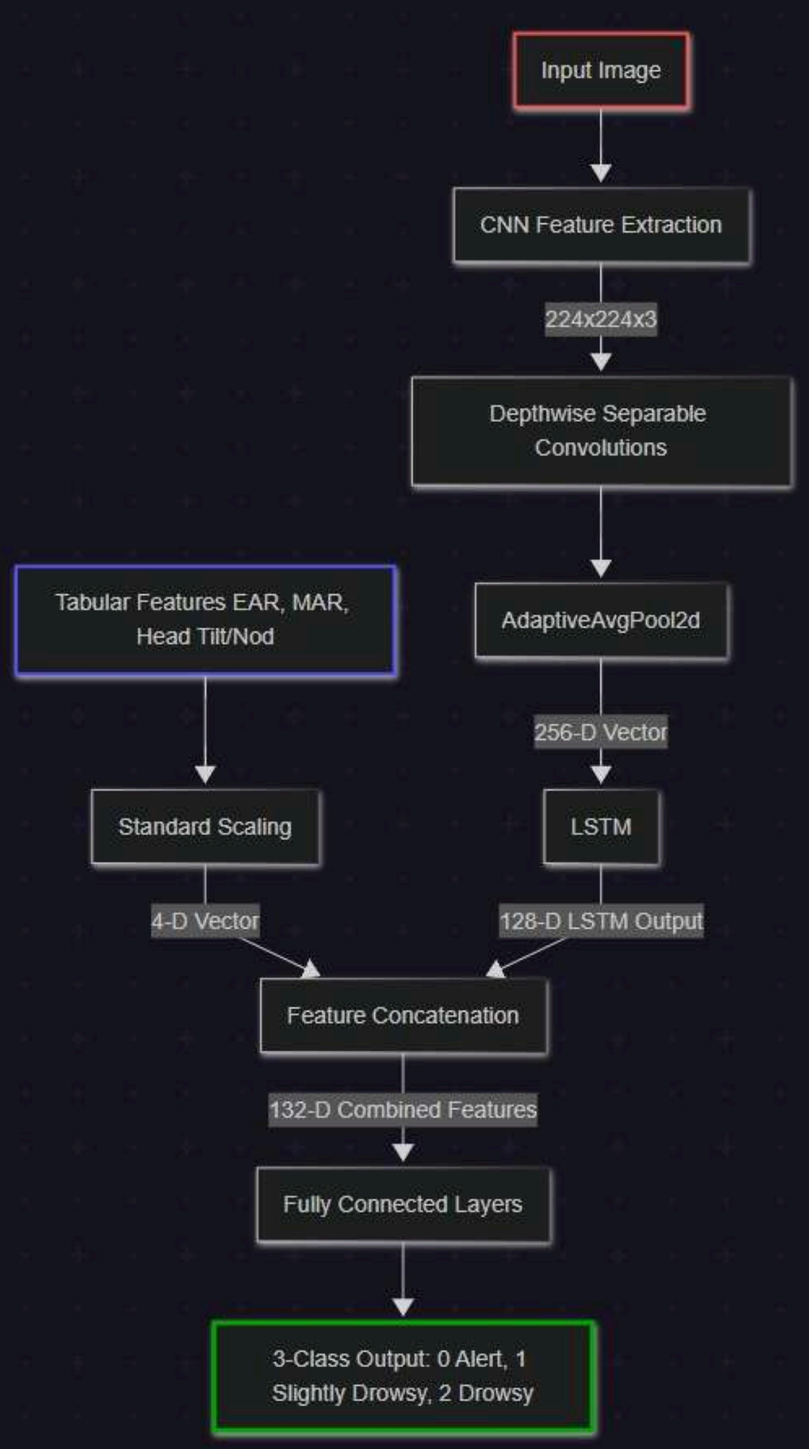
Compact Design:

1. Multi-Modal Paths
   - Visual: MobileNet-style CNN → LSTM (128 units)
   - Physiological: Processes 4 key features (EAR, MAR, Head Tilt, Nods)

2. Efficient Image Pipeline
   - Depthwise separable convolutions
   - Progressive downsampling (strided convs + adaptive pooling)

3. Fusion & Classification
   - Late fusion: LSTM (128-D) + tabular (4-D)
   - FC layers (ReLU → Dropout 30% → 3-class output)

Key Features:
✓ Facial + physiological cues ✓ Lightweight ✓ Real-time video-ready

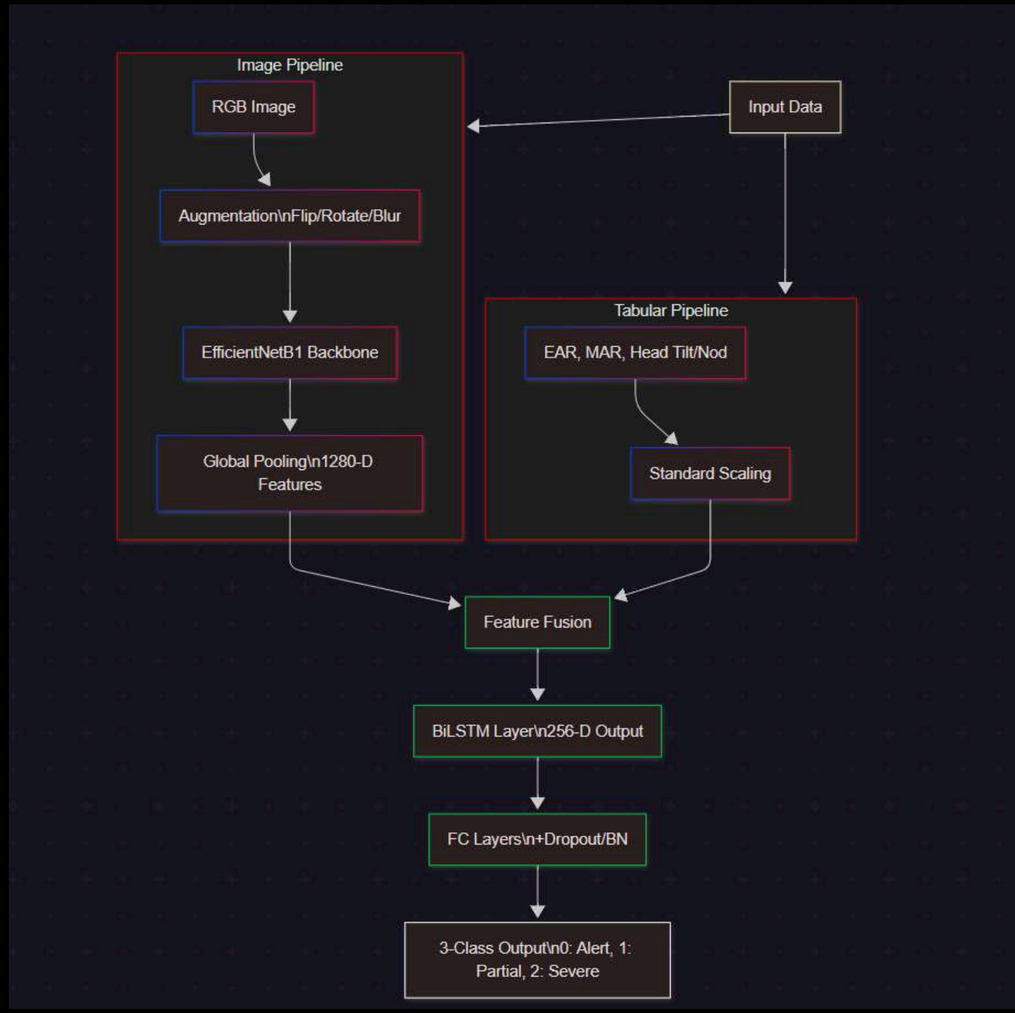


WHAT WE USED

- **CNN + LSTM + MEDIAPIPE FEATURES (EFFICIENTNET BACKBONE)**

VALIDATION ACCURACY : 85%

HIGHER INFERENCE LATENCY AND MEMORY USAGE



HYBRID MODEL ARCHITECTURE

- VISUAL PATH: **EFFICIENTNETB1** (CNN BACKBONE) → BILSTM (128 UNITS)
- PHYSIOLOGICAL PATH: STANDARDIZED TABULAR FEATURES (EAR, MAR, HEAD TILT, NOD)

OPTIMIZED TRAINING

- DATA AUGMENTATION: HORIZONTAL FLIPS, BLUR, COLOR JITTER
- REGULARIZATION: LABEL SMOOTHING (0.1), DROPOUT (40%), BATCH NORM

FUSION & DECISION

- FEATURE CONCATENATION: CNN (1280-D) + TABULAR (4-D) → BILSTM (256-D)
- CLASSIFICATION: FC LAYERS (RELU → DROPOUT → 3-CLASS OUTPUT)

KEY ADVANTAGES

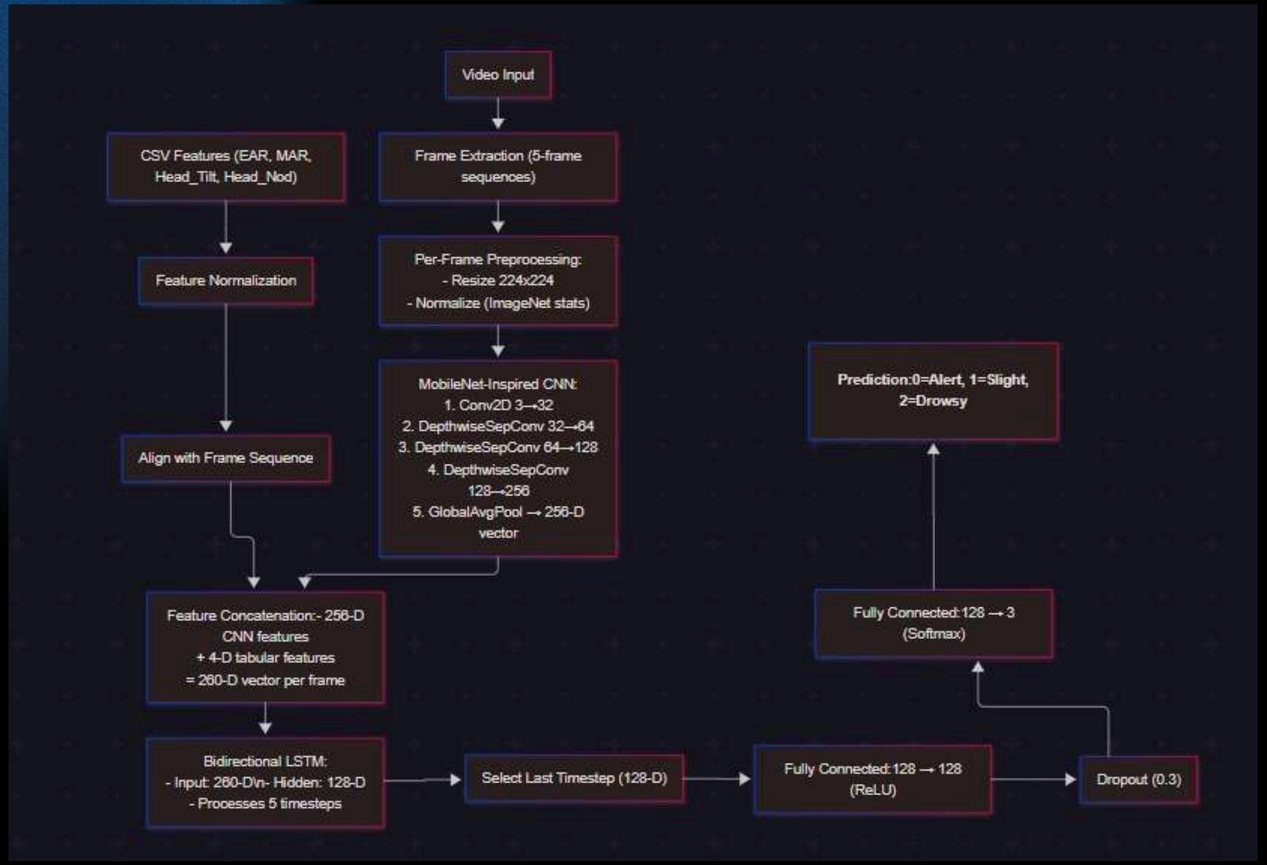
✓ MULTIMODAL (IMAGE + SENSOR FUSION)
✓ ROBUST (AUGMENTATION + BIDIRECTIONAL LSTM)
✓ DEPLOYABLE (GPU-OPTIMIZED, SAVED BEST WEIGHTS)

WHAT WE USED

- CNN + LSTM + MEDIAPIPE FEATURES

VALIDATION ACCURACY : ~90%+ (IN 3 EPOCHS)

AN LSTM WITH A SEQUENCE LENGTH OF 5 SHOWED INCREASED PROMISE BUT WE RAN OUT OF COMPUTE ON KAGGLE EVERY TIME WE TRIED IT



| PERFORMANCE METRICS

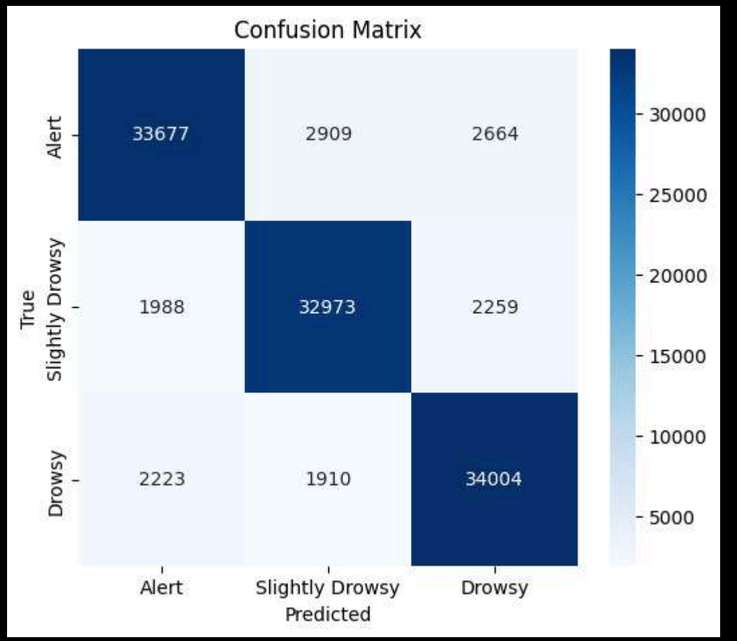
Model Size (MB) → 1.7MB

```
--- Test Set Performance ---
                precision    recall  f1-score   support

         Alert       0.89      0.86      0.87     39250
Slightly Drowsy      0.87      0.89      0.88     37220
       Drowsy        0.87      0.89      0.88     38137

      accuracy                           0.88    114607
     macro avg      0.88      0.88      0.88    114607
  weighted avg      0.88      0.88      0.88    114607

✅ Test Accuracy: 87.83%
```



Conclusion

In this project, we aimed to build an efficient and accurate drowsiness detection model that performs well even under occlusion and low-light conditions, while being lightweight enough to run on low-computational devices.

- We began with a baseline DNN, achieving 61% accuracy.
- Integrating MobileNetV2 + LSTM (1 timestep) improved accuracy significantly to 83%, balancing performance and efficiency.
- Using EfficientNet-B1 + LSTM achieved similar accuracy (~85%) but with higher memory and computational cost, making it less suitable for edge deployment.
- Finally, by using MobileNetV2 + LSTM with 5 past frames, we pushed validation accuracy to ~90%, with potential to reach 92%+ given more training epochs.

Our final approach demonstrates that with careful design, combining lightweight CNNs and temporal modeling, it's possible to build a robust, real-time drowsiness detection system that is both accurate and efficient for deployment on resource-constrained devices.

→