**Experimental Procedure:**

The experimental procedure focuses on implementing a knowledge distillation-based approach for drowsiness detection using a lightweight convolutional neural network (CNN) model. Initially, a large pre-trained teacher model, such as MobileNet with a Capsule Module, is utilized to transfer knowledge to a smaller, computationally efficient student model. The teacher model, consisting of several convolutional layers for feature extraction followed by capsule layers to capture spatial hierarchies in the data, is trained on the full dataset to achieve high performance. The primary role of the teacher model is to generate soft labels, which are probabilistic outputs that contain more nuanced information compared to hard labels. These soft labels are used to guide the training of the student model.

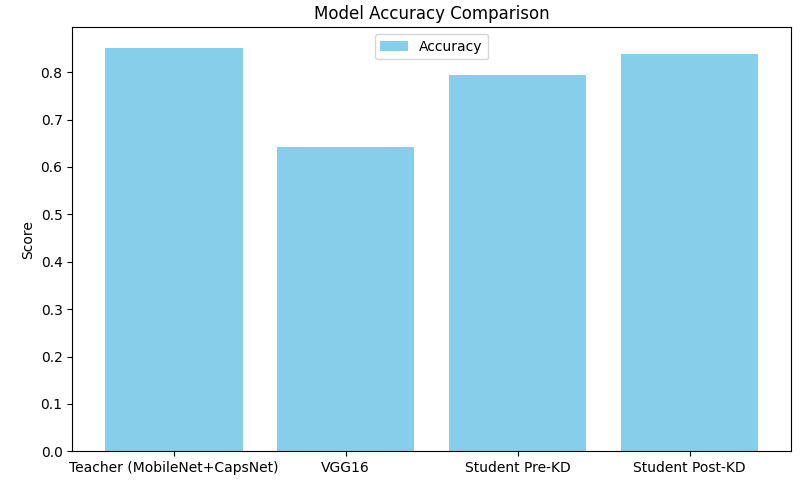
The student model is designed to be lightweight and efficient, with fewer parameters compared to the teacher model. It incorporates convolutional layers, batch normalization, ReLU activation, and pooling operations, all tailored to reduce computational requirements while retaining essential feature extraction capabilities. The final layers of the student model consist of fully connected dense layers leading to a softmax activation for classification, where it predicts whether a subject is drowsy or alert. During training, the student model is trained using both the original ground truth labels and the soft targets generated by the teacher model. The distillation process combines traditional cross-entropy loss with KL divergence loss between the outputs of the teacher and student, helping the student model to approximate the performance of the teacher model more closely.

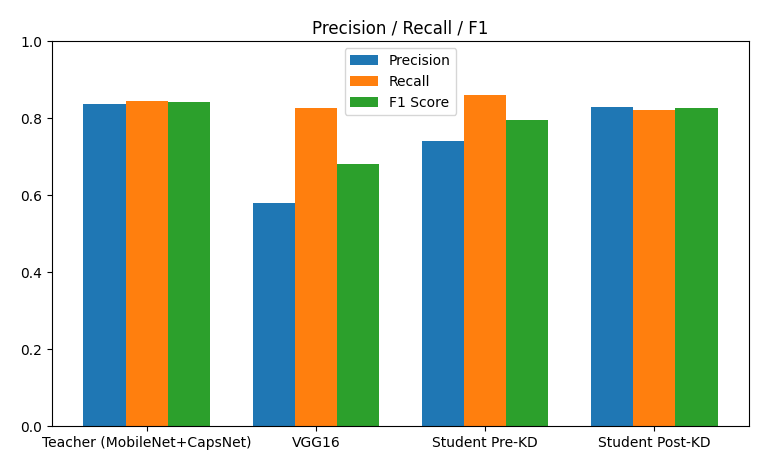
Once the knowledge distillation phase is complete, the distilled student model is integrated into a federated learning framework. In this setup, a centralized server distributes the global model to clients, each of which trains the model using its local dataset. After local training, the clients send back only the updated model weights to the server, ensuring data privacy. The server aggregates these weights and updates the global model, iterating over multiple rounds to refine the model’s performance across diverse client datasets. The performance of both the teacher model and student model is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrices, both before and after the knowledge distillation phase and after each federated learning round.

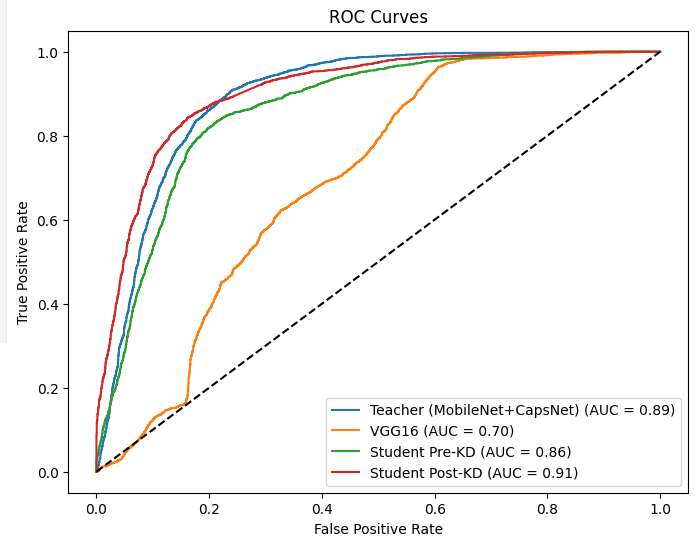
**Dataset Description:**

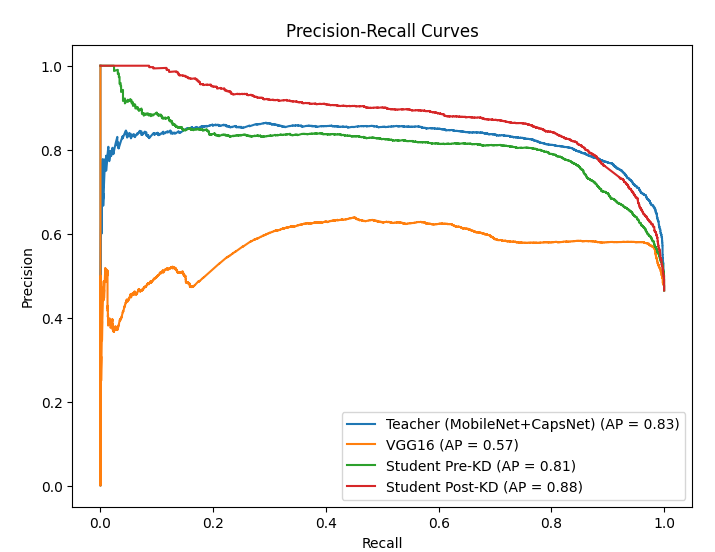
The dataset used in this project consists of 13,621 facial images specifically curated for the purpose of driver drowsiness detection. It is partitioned into 11,252 images designated for training and 2,369 images reserved for testing. The test set is carefully balanced, containing 1,223 images labeled as "Not Drowsy" and 1,151 images labeled as "Drowsy," ensuring that model evaluation remains unbiased and statistically fair. The images focus on capturing critical visual cues such as eye closure, yawning, facial expression changes, and head posture, all of which are essential indicators of drowsiness. To prevent overfitting and improve the model’s ability to generalize to unseen conditions, several data augmentation techniques are applied, including random rotations, slight scaling, horizontal flips, and controlled brightness alterations. All images are resized uniformly to 128×128 pixels to maintain a consistent input format for the models. This structured and balanced dataset, combined with robust preprocessing techniques, provides a reliable foundation for training and validating the performance of the proposed lightweight drowsiness detection system.

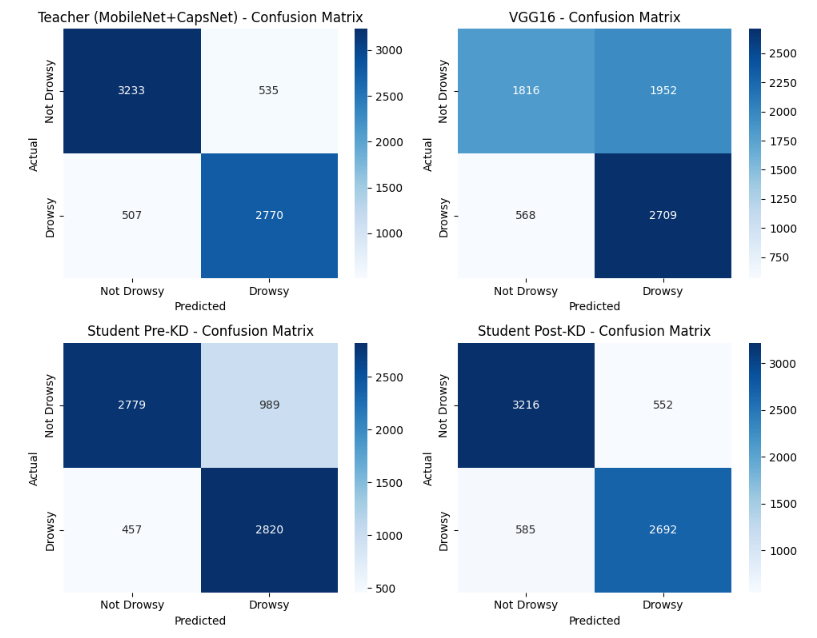
**Result:**

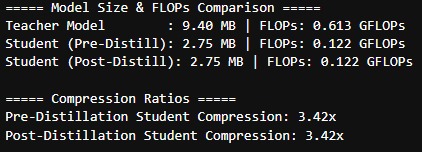
****

****

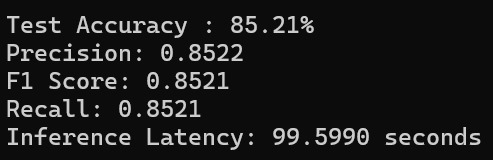
****

****

****

****

**Teacher Model** [MobileNet CapsuleNet]:

****

**Student Model** [Distilled]:

