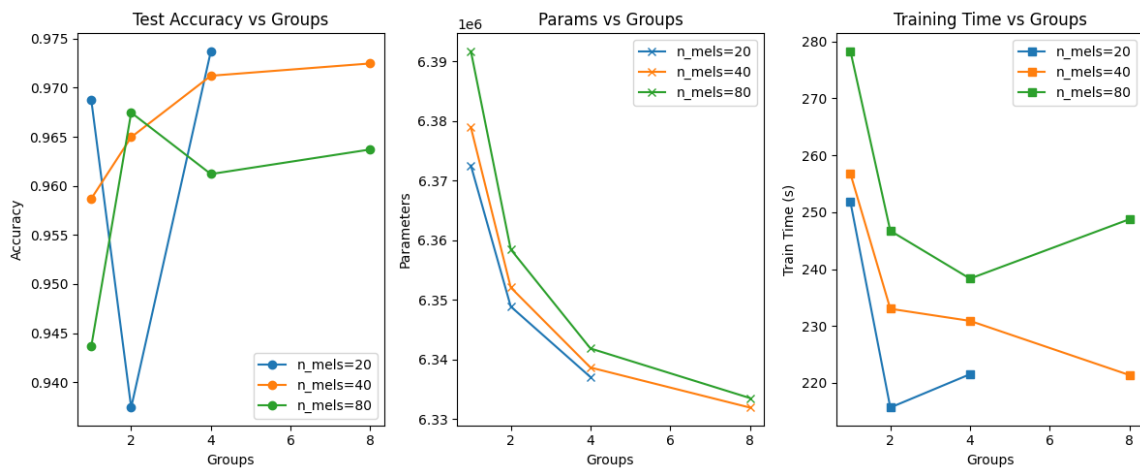


В данной работе была реализована сверточная нейронная сеть (CNN), обучаемая на аудиодатасете SpeechCommands. Целью эксперимента являлось распознавание голосовых команд 'yes' и 'no'. В качестве признакового представления использовались логарифмические мел-спектрограммы (LogMelFilterBanks).

Модель состоит из двух сверточных слоёв, за которыми следуют два полносвязных слоя. Параметры модели (число мел-коэффициентов и количество групп в свёртках) варьировались для анализа влияния на точность, время обучения и число параметров. Эксперименты проводились для значений  $n\_mels = [20, 40, 80]$  и  $groups = [1, 2, 4, 8]$ .

Результаты представлены в виде трёх графиков:

- Test Accuracy vs Groups: показывает, как меняется точность модели в зависимости от числа групп.
- Params vs Groups: демонстрирует зависимость количества параметров от группировки.
- Training Time vs Groups: отражает затраченное время обучения.



Увеличение числа групп в свёртках способствует снижению количества параметров и времени обучения, однако может негативно сказаться на точности (особенно при больших значениях  $n\_mels$ ). Оптимальное соотношение между производительностью и точностью наблюдается при  $n\_mels = 40$  и  $groups = 4$  или 8.