Model used:

The human activity recognition model we employ derives from the research presented in Hara et al.'s 2018 paper published in CVPR, titled "Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?" In this study, the authors investigate the adaptability of prevalent 2D architectures like ResNet, ResNeXt, and DenseNet, to the domain of video classification using 3D convolutional kernels.

The authors' rationale is grounded in the success of these architectures for image classification, facilitated by training on the extensive ImageNet dataset. They contend that, given the substantial scale of the Kinetics dataset, which is also sizable, these architectures can be repurposed for video classification. This involves modifying the input volume structure to encompass spatiotemporal data and incorporating 3D kernels within the architectural framework.

Data set used:

The human activity recognition model we utilize for our research is based on training with the Kinetics 400 Dataset. This dataset encompasses:

- 400 distinct classes for recognizing human activities

- A minimum of 400 video clips per class, sourced from YouTube

- A cumulative collection of 300,000 videos

For a deeper comprehension of the dataset's composition and curation process, Kay et al.'s 2017 publication titled [**The Kinetics Human Action Video Dataset**](https://arxiv.org/abs/1705.06950) offers comprehensive insights.

Reference:

<https://pyimagesearch.com/2019/11/25/human-activity-recognition-with-opencv-and-deep-learning/>