1. Get frames of pixels
2. Pass through hand detection model
3. Pass through a power convolution neural net
4. Pass through feed-forward neural net that performs classification

Mostly handled by Google MediaPipe framework which detects the landmarks

<https://github.com/kinivi/hand-gesture-recognition-mediapipe>

watch again starting around minute 45 to understand normalization