① Model → Any
② Dataset → Any
Basically Solve the problem!

Models considered: ① CNN-RNN Architecture
② VideoMAE
③ MoViNet
④ 3D CNNs

Currently the best in market according to PWC:- HERMES!
But → 8 V100 GPUs.
So No ✗

Next better VideoMAE → Uses transformers
→ Might be heavier
→ Might run out of GPU
(Lets keep it aside)

Next better CNN-RNN Architecture
→ Reliable with documentations
→ Available Colab
→ Might work let us see
→ But slightly old

Movinet → 3D CNN based but more accurate & newer
→ Super fast & efficient
→ might be the right model considering computational requirements & accuracy trade-offs

② ① → Choosing Movinet
→ If it works good.

Datasets:-

(2) ① Ucf 101 - Too small sample for read sports
- Requires preprocessing (obviously)

② HMDB 51 - No sports ⊗

③ Youtube - 8M → lot of samples
→ Unexplored datasets

④ Sports - M → Dead

→ Chose UCF-101 → With small samples
Applied preprocessing to fit DS structure

(3) ① Movinet → Too many lib issues.
Model

Flowchart →

Dataset

↓

Kinematics format

↓

Movinet-a2-stream

↓ Fine tune over 5 classes

↓

Weights

↓

TFlite → Converted due to ease of deployment

↓

Inference on MP4 → Action labels → Save for table

P.T.O