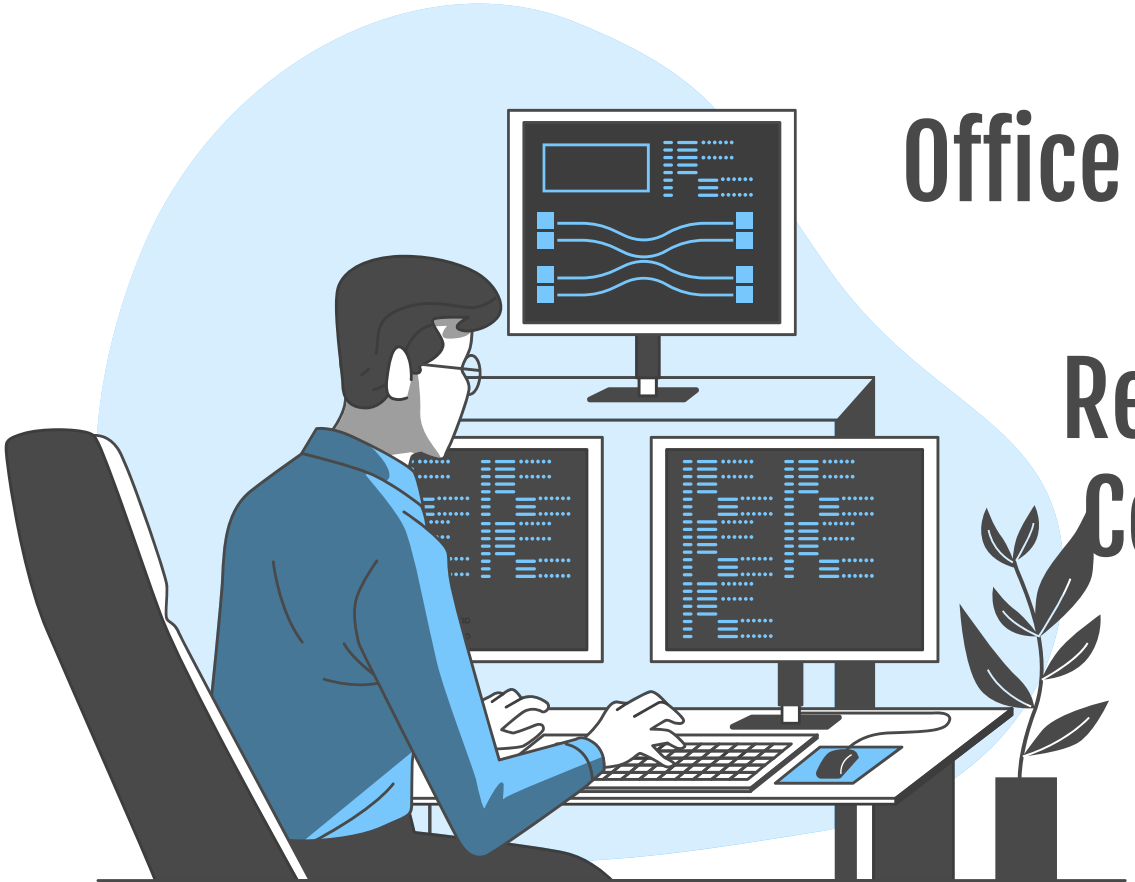


Office Working/Home Sitting Posture Recognition with Computer Vision

Capstone

Nelson Agus Kesuma
DSI-28

Table of Contents

01
Introduction

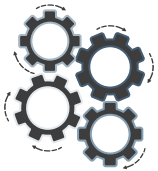
Understanding the Problem

WORK FROM HOME

WFH is more common in “new normal”

DIGITAL INTERACTION

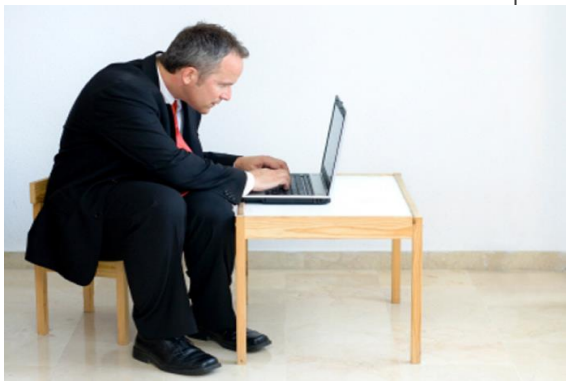
Increase in human-machine interaction

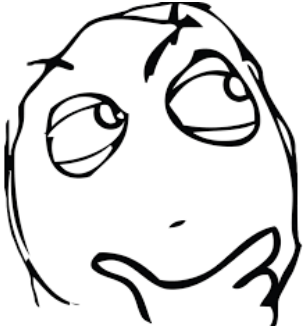


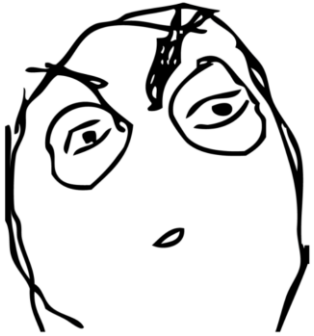
POSTURE RECOGNITION

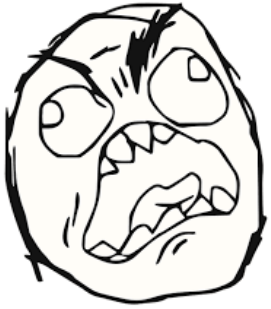
Growing needs for sitting posture recognition

Postures









Statistics At-a-Glance

One in four American adults spend more than eight hours a day sitting

On top of usual 9-5 hours, **43%** of the US population use a computer for **2** hours per day or more.
This amounted to a minimum of **10** hours per day

A survey of 8,500 participants showed that **41%** of chairs were set at wrong height (**too low**)

A survey of 8,500 participants showed that **51%** of monitors were set at unoptimized height (**too low**)

Common Effects of Bad Sitting Posture

Spine Curvature

Disrupted Digestion

Headache

Back Pain

Poor Sleep

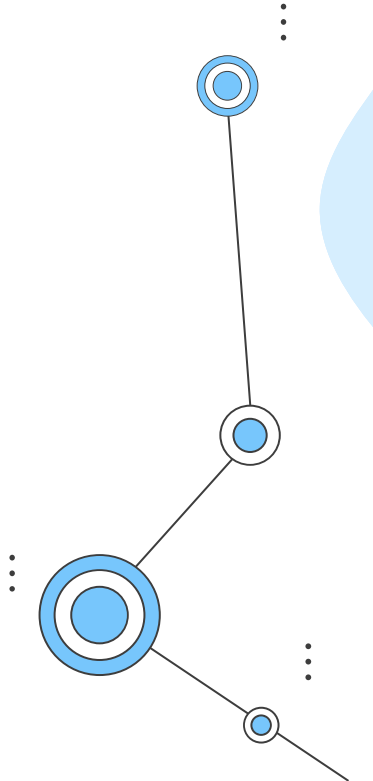
Neck Pain

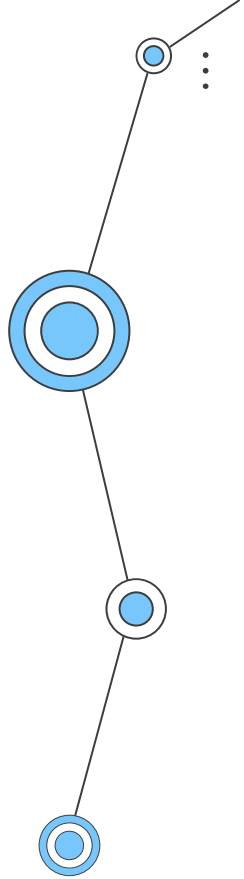
Lack of motivation

Productivity Decrease

3.5 Billions per year

Healthcare cost estimation by Workplace Safety and Health (WSH) Singapore





Propose to develop Computer Vision live posture classification

There are myriad cost and effects due to unmonitored bad sitting posture during work

Develop Machine Learning Model to accurately predict multiclass sitting posture classification

Evaluation Criteria:

- High test accuracy score of above 90% on validation set
- High F1-score of above 90% on validation set

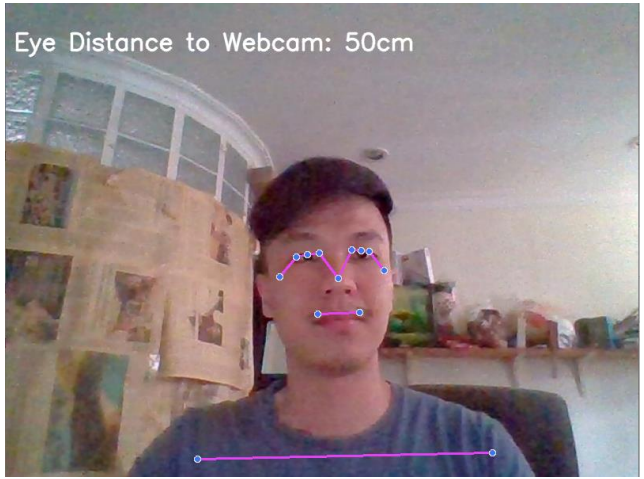
02

Dataset & Exploratory Data Analysis

Initial Dataset Overview

Webcam

- Taken ownself through webcam
- Consist of 33 feature landmarks (xyz and v coordinates)



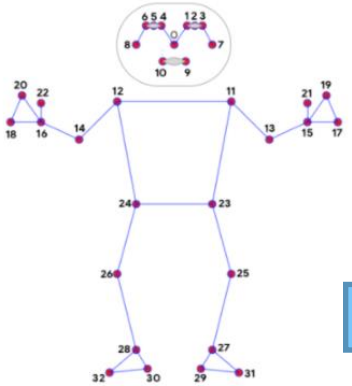
FEATURES	DESCRIPTION
133 columns	Posture
1720 rows	132 coordinates (4 * 33 landmarks)

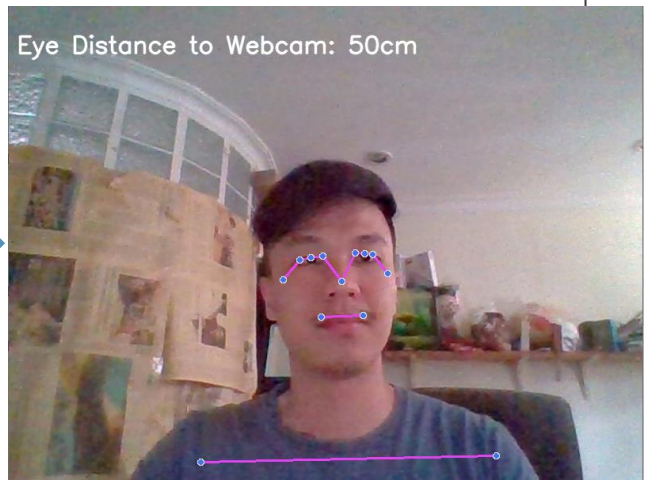
Tools to Get There

MediaPipe

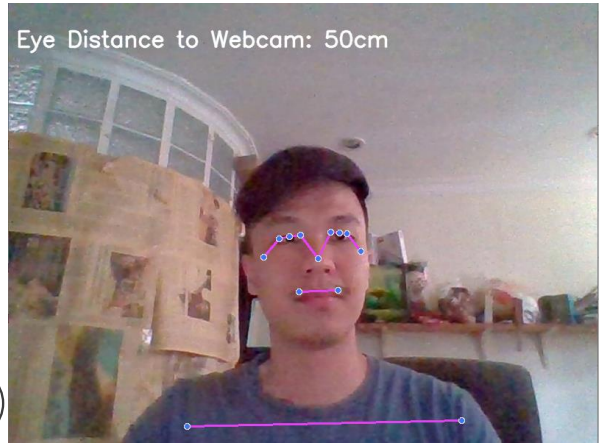
- Google Open Source framework
- Accurate Pose Estimation for media processing
- Cross-platform friendly. It runs on Android, iOS, web and Youtube servers
- Combination of Face mesh, Pose classification, object detection, selfie segmentation etc.

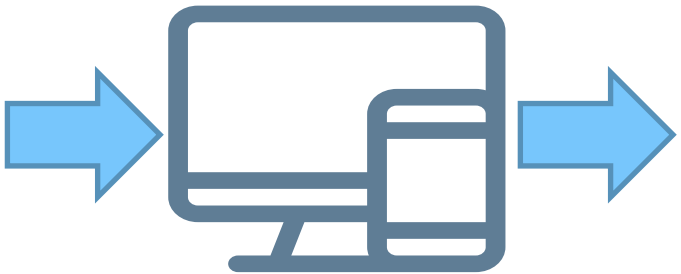
Tools to Get There

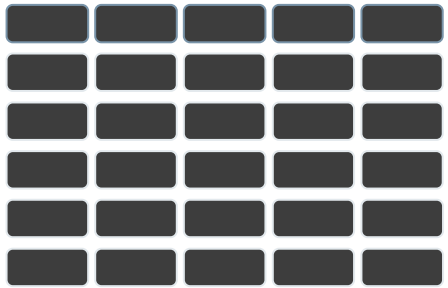




Extract & Export







Record Body Landmarks

Convert Landmarks to Coordinates

Export as CSV

Feature Engineering

Eye Distance to Camera

- Important to determine eye closeness to screen
- When people slouch, their eyes are usually closer to the screen
- Ideal distance to screen is usually one's arm length

The Posture Classification

Straight

Slouched

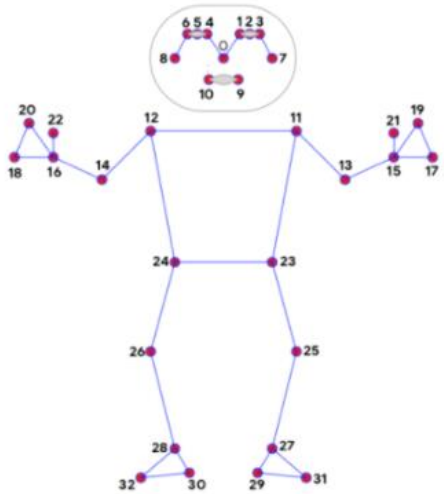
Slouched + Roundshoulder

Lieback

Uneven shoulder

The Webcam Dataset

X1: x-coordinate of nose
Y1: y-coordinate of nose
Z1: z-coordinate of nose
V1: Likelihood of nose in the frame



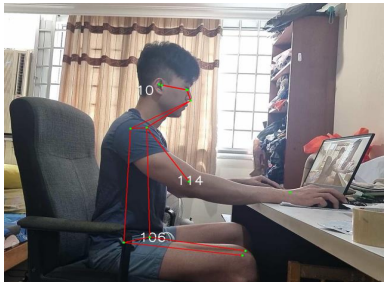
0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index

	posture	distance	x1	y1	z1	v1	x2	y2	z2	v2	x3	y3	z3	v3	x4	y4	z4	v4	x5	y5	z5	v5
0	straight	54.482725	0.540115	0.643918	-1.082978	0.999611	0.560545	0.597951	-0.987799	0.999653	0.574086	0.600362	-0.988065	0.999710	0.587434	0.603645	-0.988064	0.999675	0.515215	0.599925	-0.991564	0.999585
1	straight	52.867762	0.545043	0.646108	-1.093874	0.999447	0.562441	0.598812	-1.000388	0.999513	0.575351	0.601176	-1.000781	0.999584	0.588243	0.604396	-1.000806	0.999543	0.517219	0.600766	-1.003278	0.999428
2	straight	53.642567	0.544261	0.646108	-1.052980	0.999187	0.561833	0.599122	-0.964017	0.999303	0.575026	0.601721	-0.964510	0.999398	0.587763	0.605226	-0.964558	0.999354	0.515784	0.600763	-0.960265	0.999193
3	straight	52.867762	0.544072	0.646125	-1.001081	0.999030	0.561759	0.599375	-0.912840	0.999171	0.575022	0.602080	-0.913346	0.999263	0.587767	0.605714	-0.913323	0.999211	0.515045	0.600840	-0.912368	0.999065
4	straight	53.683076	0.542942	0.649676	-1.015812	0.998835	0.561096	0.602478	-0.920225	0.998980	0.574674	0.605388	-0.920551	0.999103	0.587378	0.609061	-0.920404	0.999041	0.513867	0.603527	-0.927293	0.998854

Initial Dataset Overview

Sidecam

- Taken ownself through sidecam (left, right, diagonal right, diagonal left)
- Consist of 17 feature landmarks (xy-coordinates and s)



FEATURES	DESCRIPTION
52 columns	Posture
5498 rows	51 coordinates (3 * 17 landmarks)

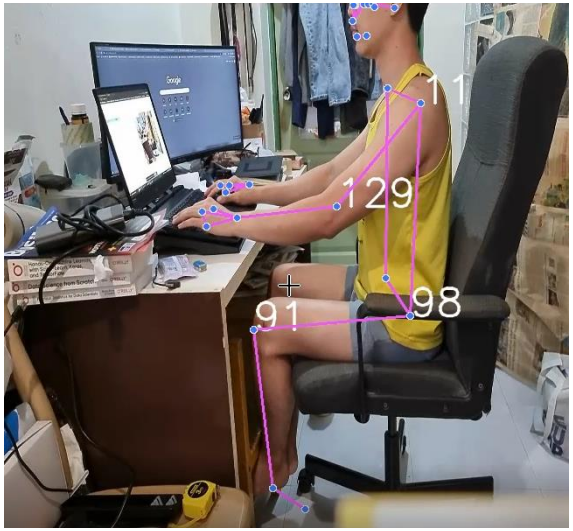
Tools to Get There

TensorFlow MoveNet

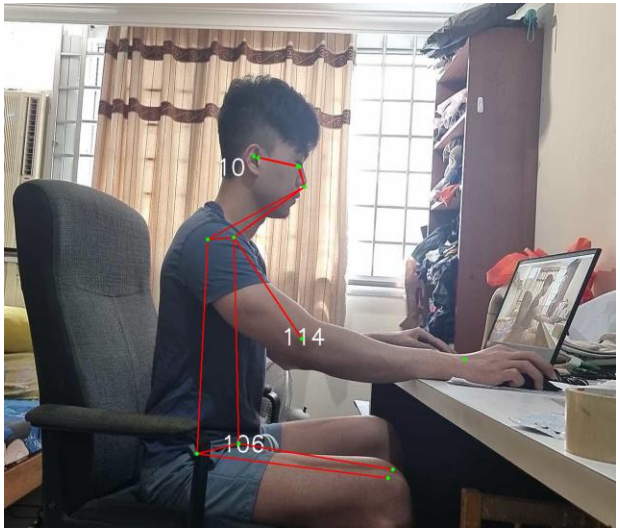
- Google Open Source framework from Tensorflow
- Latest release in 2021 of Accurate Pose Estimation for media processing library
- Good prediction for covered joint areas

The Difference

MediaPipe



Tensorflow Movenet



MediaPipe requires an initial pose alignment and dataset to where either the whole person is visible or where hips and shoulders keypoints can be confidently annotated.

Feature Engineering

Neck Inclination

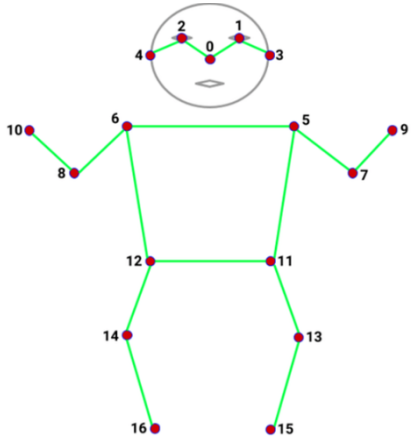
- When people hunched, neck inclination angle is larger
- Measured by using cross product of 2 vectors between shoulder (P1), ear (P2) and imaginary point (P3)

Left Hip Angle, Left Bicep Angle, Right Hip Angle, Right Bicep Angle

- Key angles for determining posture differences

The Sidecam Dataset

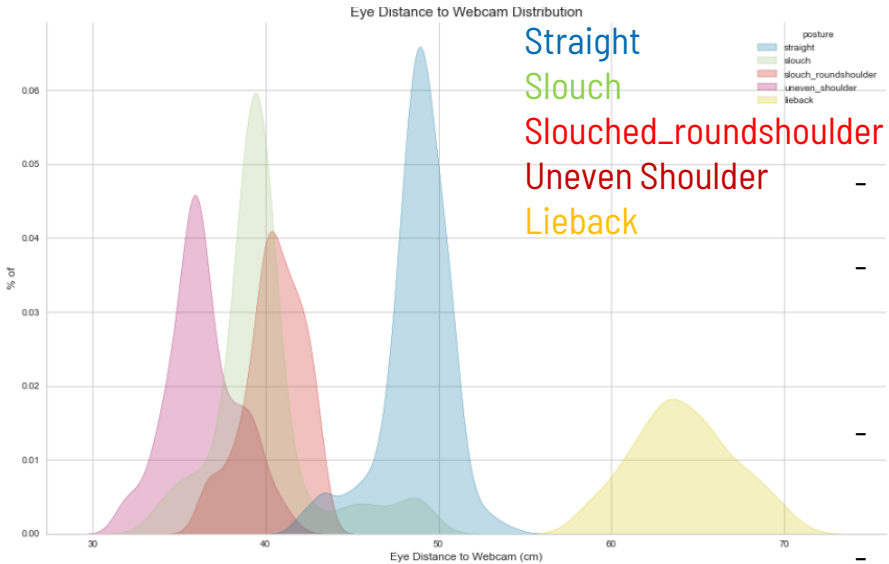
X1: x-coordinate of nose
Y1: y-coordinate of nose
S1: Confidence score of nose in the frame



0. Nose
1. Left Eye
2. Right Eye
3. Left Ear
4. Right Ear
5. Left Shoulder
6. Right Shoulder
7. Left Elbow
8. Right Elbow
9. Left Wrist
10. Right Wrist
11. Left Hip
12. Right Hip
13. Left Knee
14. Right Knee
15. Left Ankle
16. Right Ankle

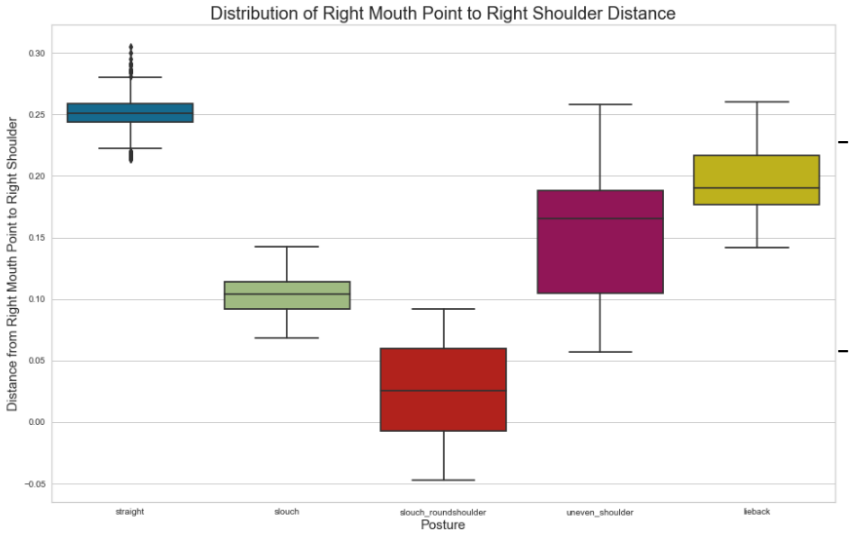
	class	neck_inclination	left_hip_angle	left_bicep_angle	right_hip_angle	right_bicep_angle	y1	x1	s1	y2	...	s14	y15	x15	s15	y16	x16	s16	y17	x17	s17
0	straight	18.148982	88.293295	110.395569	82.380660	98.037127	0.209674	0.585207	0.642684	0.163057	...	0.129741	0.776018	0.430507	0.509525	0.906293	0.482469	0.026621	0.853896	0.569948	0.073047
1	straight	16.296187	112.163444	113.258047	86.573638	98.215743	0.198345	0.590137	0.615752	0.156344	...	0.096807	0.779168	0.430096	0.473967	0.915161	0.454808	0.032801	0.854644	0.569576	0.060822
2	straight	15.440235	110.157745	112.501603	86.183511	100.264903	0.198931	0.594332	0.633484	0.154357	...	0.112910	0.777814	0.429936	0.462543	0.911040	0.455521	0.036156	0.857120	0.584999	0.075562
3	straight	14.636403	113.576665	114.833703	86.478473	94.080118	0.191974	0.597751	0.563277	0.146848	...	0.060127	0.774932	0.429564	0.551982	0.920200	0.437648	0.031089	0.879962	0.535686	0.045040
4	straight	14.121881	111.263034	113.958558	87.001686	99.058500	0.182014	0.598116	0.651450	0.137654	...	0.103613	0.775958	0.429372	0.552059	0.910851	0.485200	0.025875	0.881896	0.554634	0.059867

Eye closeness to screen distribution (Webcam)



- Eye Distance to Webcam distribution differs from each posture
- For straight posture, data is more distributed around 50cm which is the average human hand length. Thus, this is the correct sitting posture distance
- For slouch, slouch_roundshoulder and uneven_shoulder, data centers around a closer distance to webcam
- For lieback posture, eye distance to camera is usually further away from the webcam

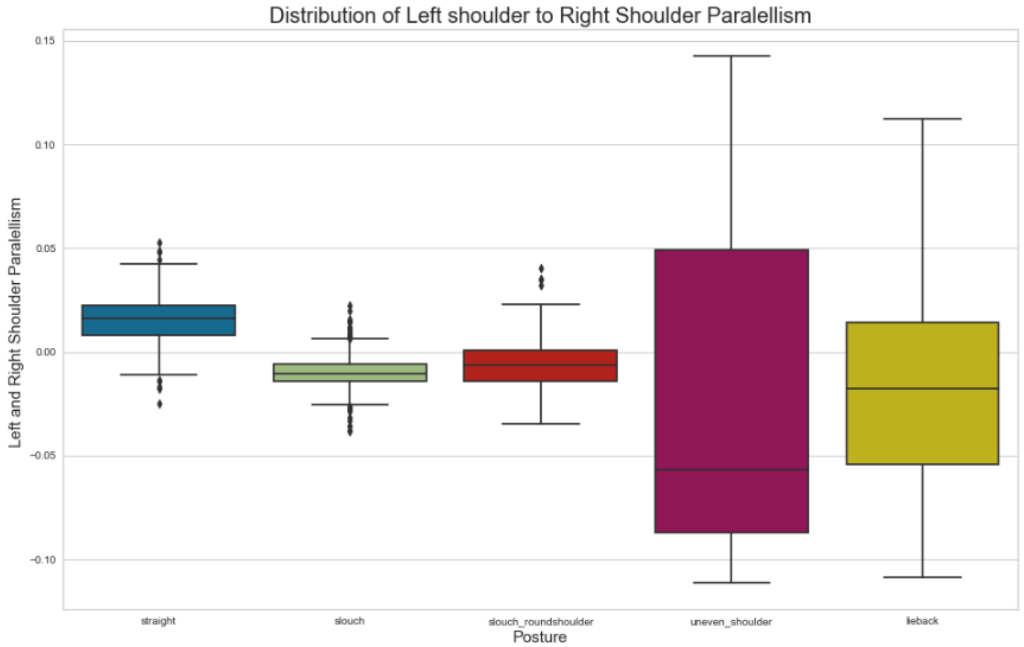
Mouth Distance to Shoulder by Postures (Webcam)



- There is a trend observed from straight posture to slouch_roundshoulder posture. This is due to the nature of slouching in which the mouth is closer to shoulder and it is even closer when people have roundshoulder

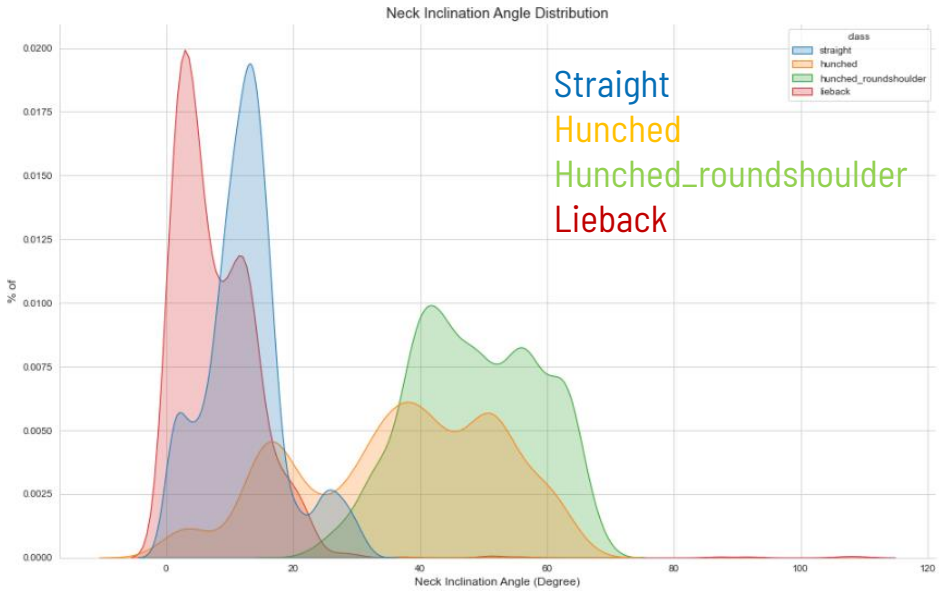
- For uneven shoulder and lieback posture, mouth closeness to shoulder is not as severe as slouching. However, noted that distance is still closer as compared to a straight posture

Shoulder Parallelism (Webcam)



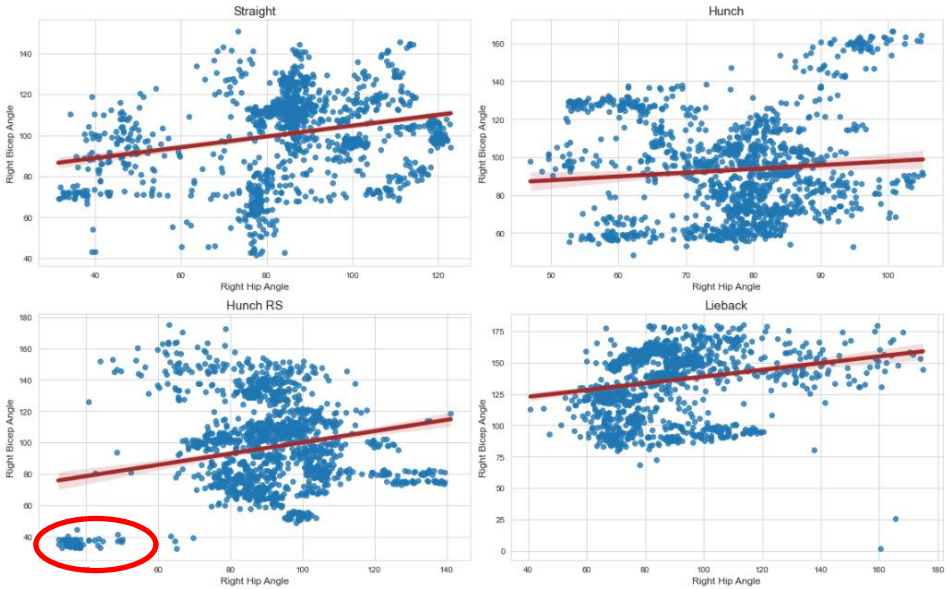
For uneven shoulder, shoulder is indeed not parallel while other postures are relatively parallel

Neck Inclination Angle Distribution (Sidecam)



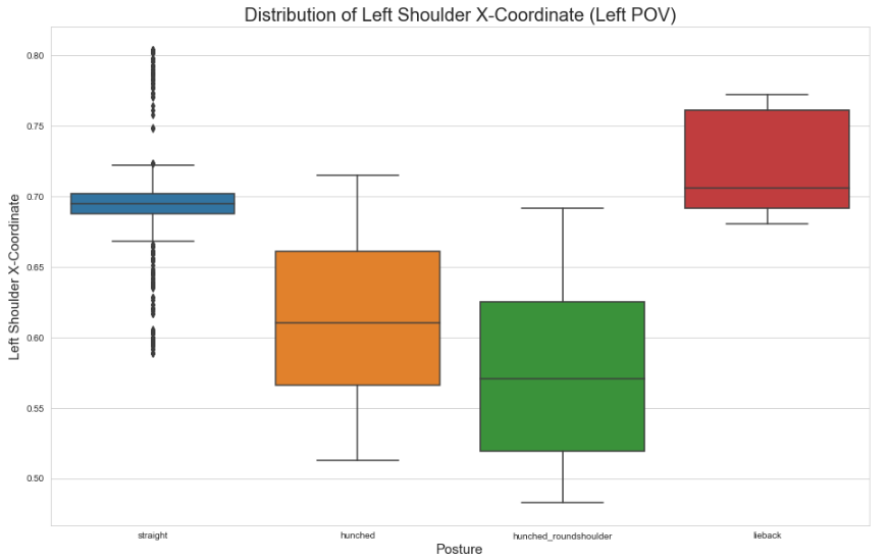
- Straight and lieback posture neck inclination angle is the lowest
- Hunched and hunched_roundshoulder postures have the largest neck inclination angle

Arm and Torso Angle Correlation (Sidecam)



- Generally all postures have similar correlation
- Hunched_roundshoulder has datapoints trained at very low bicep angle and hip angle
- Lieback has datapoints trained at higher bicep angle

Shoulder relativity to frame (Left POV) (Sidecam)



Hunched_roundshoulder has a more forward shoulder than other posture

03

Modelling & Model Evaluation

Modelling

Logistic Regression

Fits data on a sigmoid curve to classify 2 categories

Ridge Classifier

Similar to ridge regression, only converts target data into [-1,1] and solve with ridge regression method

Gradient Boosting Classifier

Combines weak classifiers into a strong classifier by taking a subset of data in each iteration and learning through mistakes

Random Forest Classifier

Ensemble of decision trees to vote on the predicted class. Predicted class is decided through majority outcome from each decision tree

Adaptive Boosting Classifier (ADA)

Combines weak classifiers into a strong classifier by learning through mistakes

Decision Tree Classifier

Single Decision tree outcome prediction

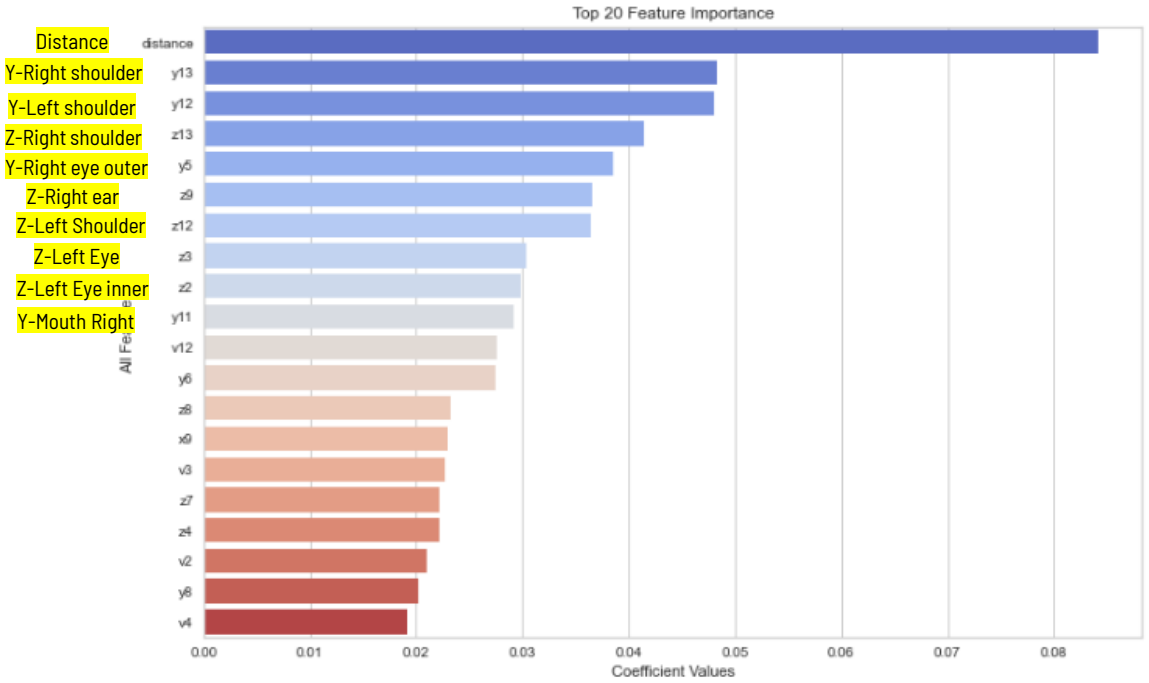
Modelling

Webcam

Models	Validation Accuracy	Mean F1 Score of 5 Postures
Logistic Regression	99.8%	99.8%
Ridge Classification	97.2%	97.2%
Random Forest Classifier	99.6%	99.6%
Gradient Boosting Classifier	99.6%	99.5%
Adaptive Boosting (ADA) Classifier	53.2%	45.1%
Decision Tree Classifier	98.6%	98.6%

Logistic Regression, Random Forest and Gradient Boosting Classifier have relatively high validation accuracy score and F1 score

Feature importance to classify a posture (Webcam)



Above is the top 3 key predictors for the model to classify a posture:

- Distance
- y13 (Right Shoulder)
- y12 (Left Shoulder)

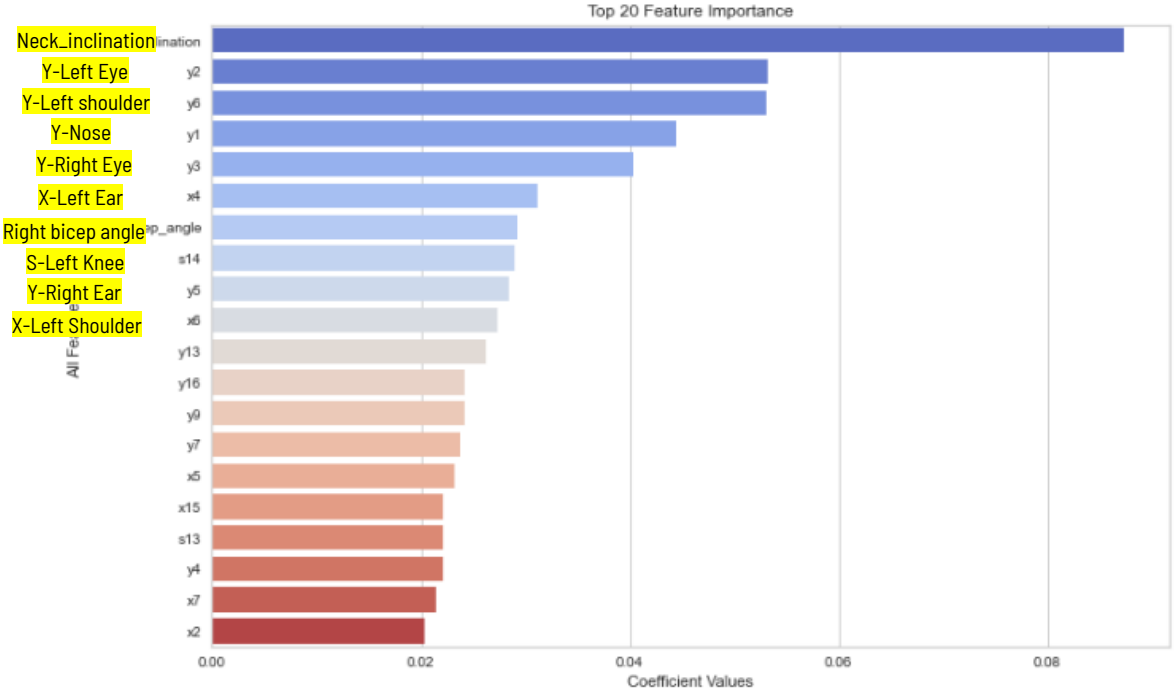
Modelling

Sidecam

Models	Validation Accuracy	Mean F1 Score of 5 Postures
Logistic Regression	98.8%	98.8%
Ridge Classification	95.2%	94.5%
Random Forest Classifier	99.6%	99.1%
Gradient Boosting Classifier	99.5%	99%
Adaptive Boosting (ADA) Classifier	59.4%	57.7%
Decision Tree Classifier	97.4%	97%

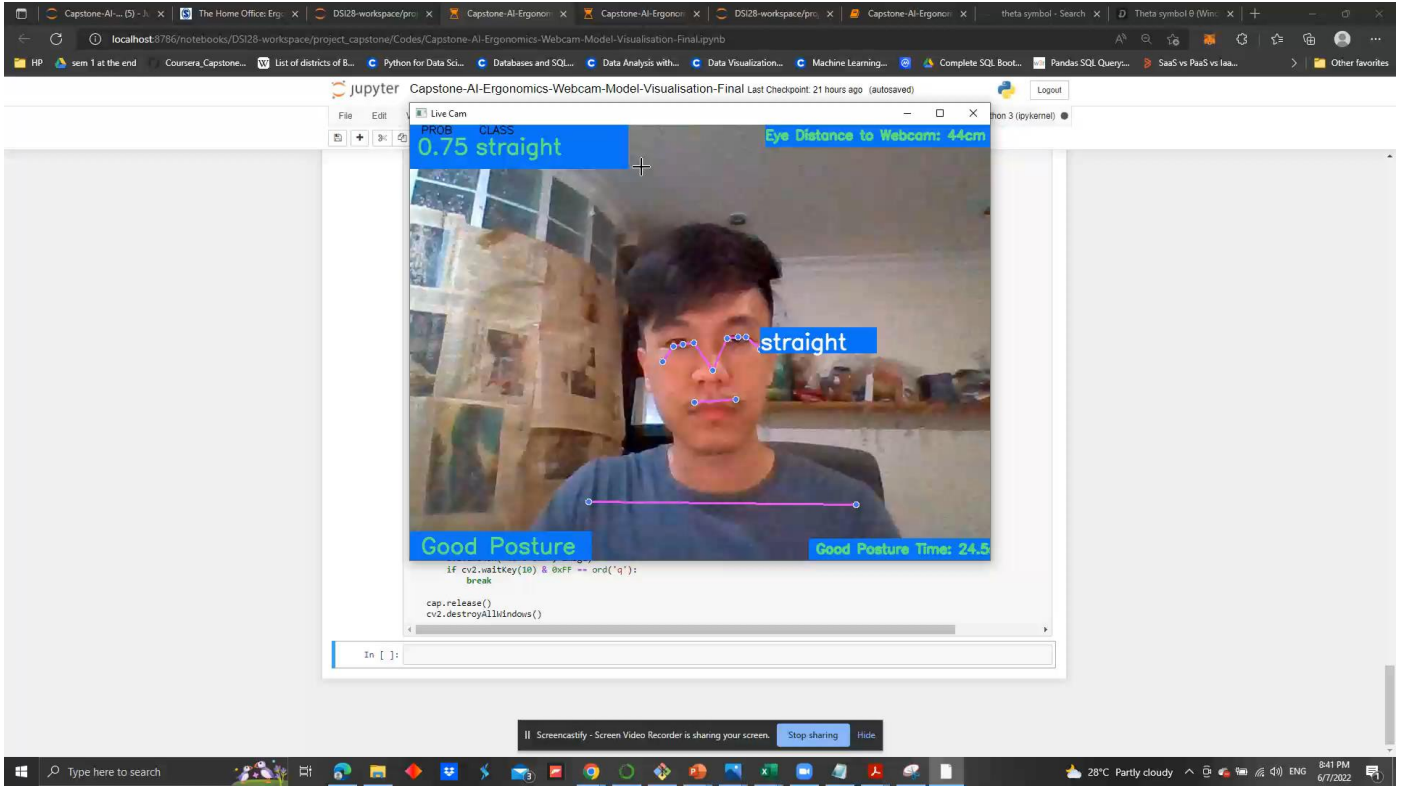
Logistic Regression, Random Forest and Gradient Boosting Classifier have relatively high validation accuracy score and F1 score

Feature importance to classify a posture (Sidecam)

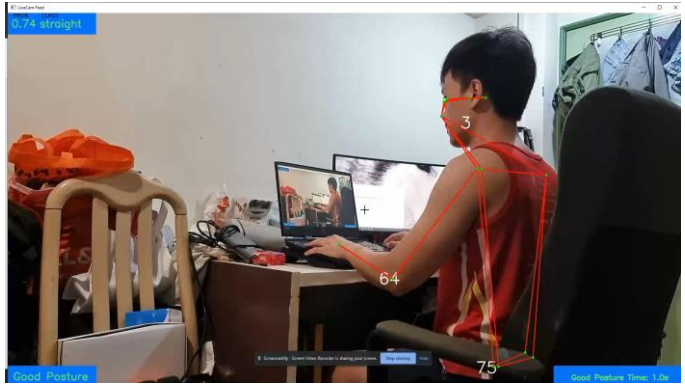


Above is the top 3 key predictors for the model to classify a posture:
- Neck inclination
- y2 (Left Eye)
- Y6 (Left Shoulder)

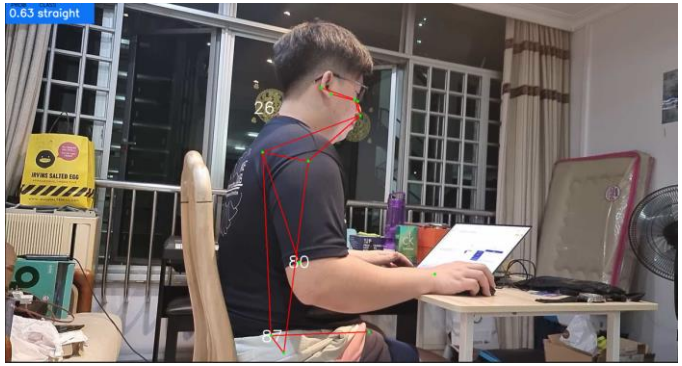
Webcam Video Prediction



Sidecam Video Prediction



Sidecam Video Prediction



04

Live Demo

05

Conclusions & Next Steps

Conclusions

Best Model: **Random Forest Classifier**

- Achieved 99.6% validation score on both webcam & side cam
- Achieved ~99% F1 score on both webcam & side cam
- Computer Vision Live Prediction score can be further enhanced with more training datasets through different camera angles and camera placements
- However, model is able to capture postures relatively well and are able to capture landmarks & features

Challenges/Limitations

- Models are only able to predict straight, slouched/hunched, slouched/hunched + roundshoulder, uneven shoulder and lieback.
- In real life, there are other postures affecting ergonomics (cross legs, butterfly sit posture, knee angle detection, etc)
- Models are too sensitive to camera angle and placements. Thus, moving the camera to train improved the computer vision accuracy. However, more coverage area is needed for training.

Next Steps

- More complex postures specifically legs joints as these body parts also affect sitting ergonomics.
- Utilize more cameras to cover all camera angles/placements for model training. Thus, it will be able to create a much robust model at different placements when deployed to offices/homes.
- Train more models to cater for 2 or more people in a frame.
- Explore deep learning models (CNN or RNN) for possibly improve predictive power.
- Use of object detection models such as Mask R-CNN or YOLO to detect monitors, chairs and tables to ultimately give recommendations for best alignments/placements and heights for each workstation.
- Use Tkinter to develop GUI app for posture record and recommendations to correct postures.

Thanks!

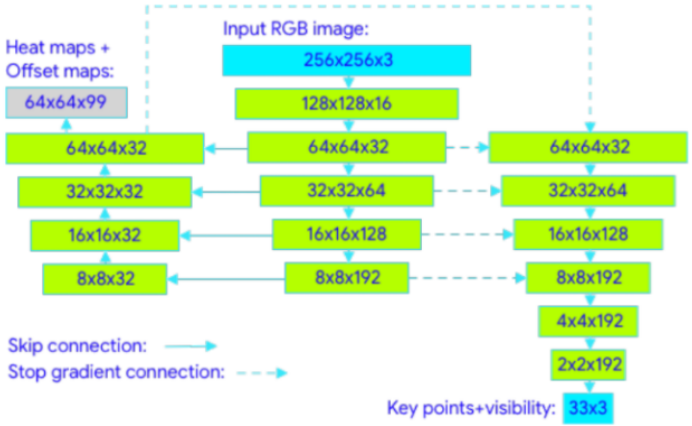
Do you have any questions?

Appendix

MediaPipe vs MoveNet

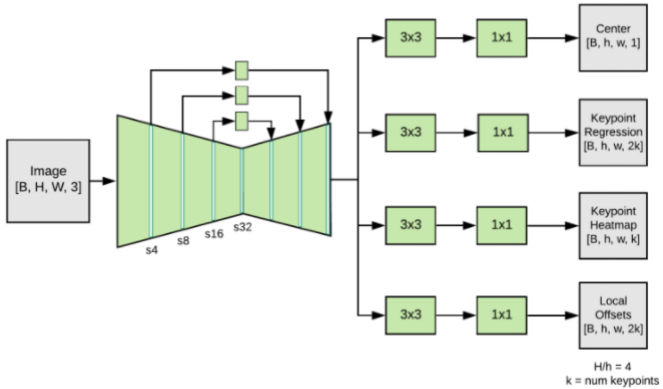
MediaPipe

Top-Down Models



MoveNet

Bottom-Up Models



MoveNet: Architecture Based on 4 Predictor Heads

MediaPipe vs MoveNet

MediaPipe

The input to the model is 256X256X3 which gets reduced to 8X8X192 through a bunch of convolution blocks then the encoder output is upsampled along with skip connections.
The output of the encoder-decoder is 64X64X32 which is passed to the coordinate regression network and another convolution layer that produces heatmap and offset map of output 64X64X99.
Since there are 33 key points we can say for every keypoint it predicts 3 images.
Coordinate regression network predicts the key points and visibility of key points. The network itself has backpropagation but has a stop gradient connection with the encoder.
Intuitively, we can say that the coordinate regression network has no control over the encoder and decoder but leverages its information for predicting keypoints.
The output of the Coordinate regression network is 33X3 which is for each keypoints we have x_coordinate, y_coordinate, and visibility.

MoveNet

Four prediction heads are attached to the feature extractor:
1. Person center heatmap to predict geometric center of person instances
2. Keypoint regression field to predict full set of keypoints for a person to group keypoints into instances.
3. Person keypoint heatmap to predict location of all keypoints irrespective of person instances
4. 2D per-keypoint offset field to predict local offsets from each output feature map pixel to the precise sub-pixel location of each keypoint.

PyCaret Model

Webcam

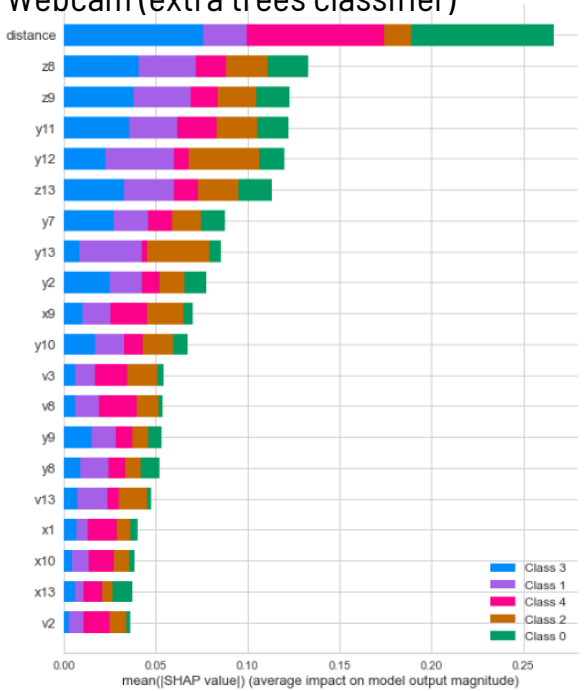
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
qda	Quadratic Discriminant Analysis	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0050
et	Extra Trees Classifier	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0490
rf	Random Forest Classifier	0.9983	1.0000	0.9984	0.9984	0.9983	0.9979	0.9979	0.0770
gbc	Gradient Boosting Classifier	0.9975	1.0000	0.9975	0.9976	0.9975	0.9969	0.9969	0.6540
lr	Logistic Regression	0.9967	1.0000	0.9968	0.9968	0.9967	0.9958	0.9958	0.6950
svm	SVM - Linear Kernel	0.9967	0.0000	0.9966	0.9968	0.9966	0.9958	0.9958	0.0070
lda	Linear Discriminant Analysis	0.9967	1.0000	0.9970	0.9969	0.9967	0.9958	0.9959	0.0040
lightgbm	Light Gradient Boosting Machine	0.9967	1.0000	0.9968	0.9969	0.9967	0.9958	0.9959	0.4520
xgboost	Extreme Gradient Boosting	0.9959	1.0000	0.9961	0.9961	0.9959	0.9948	0.9948	0.1950
knn	K Neighbors Classifier	0.9917	1.0000	0.9911	0.9921	0.9916	0.9895	0.9896	0.2390
dt	Decision Tree Classifier	0.9909	0.9942	0.9904	0.9914	0.9907	0.9885	0.9887	0.0060
ridge	Ridge Classifier	0.9668	0.0000	0.9707	0.9703	0.9668	0.9581	0.9591	0.0040
nb	Naive Bayes	0.9385	0.9937	0.9424	0.9405	0.9381	0.9224	0.9231	0.0040
ada	Ada Boost Classifier	0.7024	0.8656	0.6937	0.5996	0.6239	0.6213	0.6736	0.0460
dummy	Dummy Classifier	0.2544	0.5000	0.2000	0.0647	0.1032	0.0000	0.0000	0.0040

Sidecam

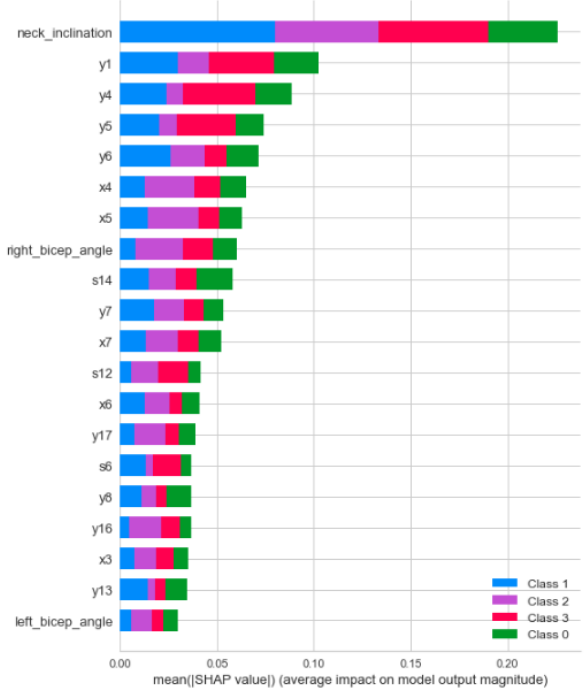
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
et	Extra Trees Classifier	0.9992	1.0000	0.9993	0.9992	0.9992	0.9990	0.9990	0.0820
lightgbm	Light Gradient Boosting Machine	0.9971	1.0000	0.9973	0.9972	0.9971	0.9962	0.9962	0.8090
rf	Random Forest Classifier	0.9966	1.0000	0.9968	0.9967	0.9966	0.9955	0.9955	0.2210
knn	K Neighbors Classifier	0.9964	0.9998	0.9966	0.9964	0.9964	0.9951	0.9952	0.2550
xgboost	Extreme Gradient Boosting	0.9961	1.0000	0.9962	0.9962	0.9961	0.9948	0.9948	1.0850
gbc	Gradient Boosting Classifier	0.9953	0.9999	0.9955	0.9954	0.9953	0.9937	0.9938	3.2200
qda	Quadratic Discriminant Analysis	0.9940	0.9999	0.9941	0.9941	0.9940	0.9920	0.9920	0.0110
lr	Logistic Regression	0.9891	0.9994	0.9894	0.9893	0.9891	0.9854	0.9855	0.7140
svm	SVM - Linear Kernel	0.9808	0.0000	0.9809	0.9810	0.9808	0.9743	0.9743	0.0170
dt	Decision Tree Classifier	0.9769	0.9845	0.9768	0.9773	0.9768	0.9691	0.9693	0.0270
lda	Linear Discriminant Analysis	0.9673	0.9976	0.9675	0.9679	0.9672	0.9562	0.9565	0.0140
ridge	Ridge Classifier	0.9527	0.0000	0.9517	0.9532	0.9523	0.9367	0.9372	0.0060
nb	Naive Bayes	0.8454	0.9698	0.8471	0.8493	0.8465	0.7933	0.7938	0.0070
ada	Ada Boost Classifier	0.6640	0.8693	0.6606	0.6856	0.6548	0.5503	0.5603	0.1870
dummy	Dummy Classifier	0.2817	0.5000	0.2500	0.0794	0.1238	0.0000	0.0000	0.0070

PyCaret Model

Webcam (extra trees classifier)

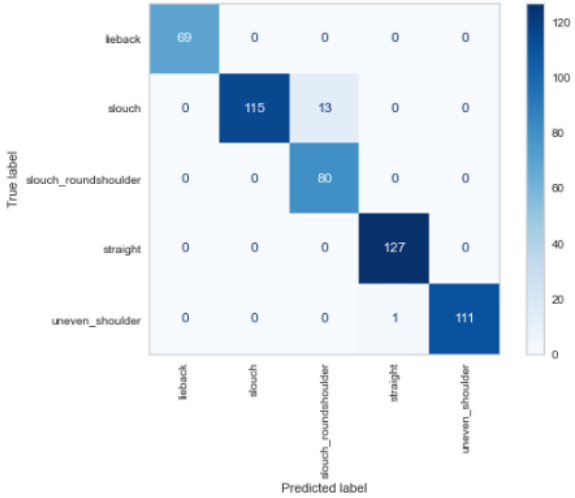


Sidecam (extra trees classifier)



Confusion matrix (Random Forest)

Webcam



Sidecam