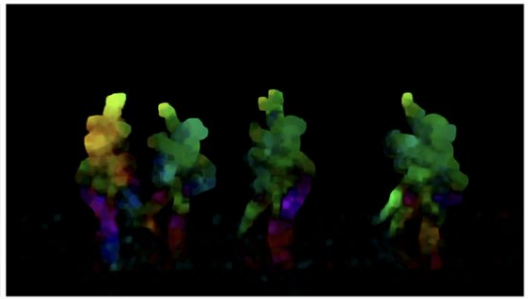
Blake Bleier
Shruti Jain
Summer McGrogan
Danny Nguyen

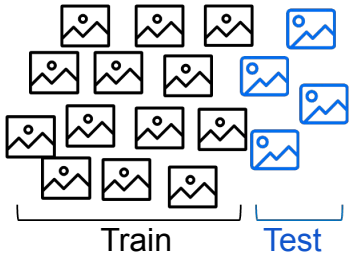
What and why?

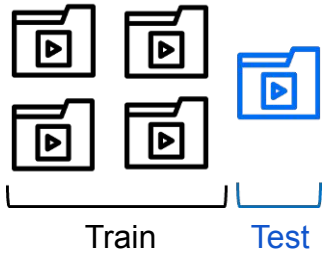
Dataset

- Dance videos split into frames
- Divided into 3 partitions
  - Original frames
  - Optical flow
  - Skeletal (visual & JSON)
- Started with dance move generation, simplified into classification due to time constraints



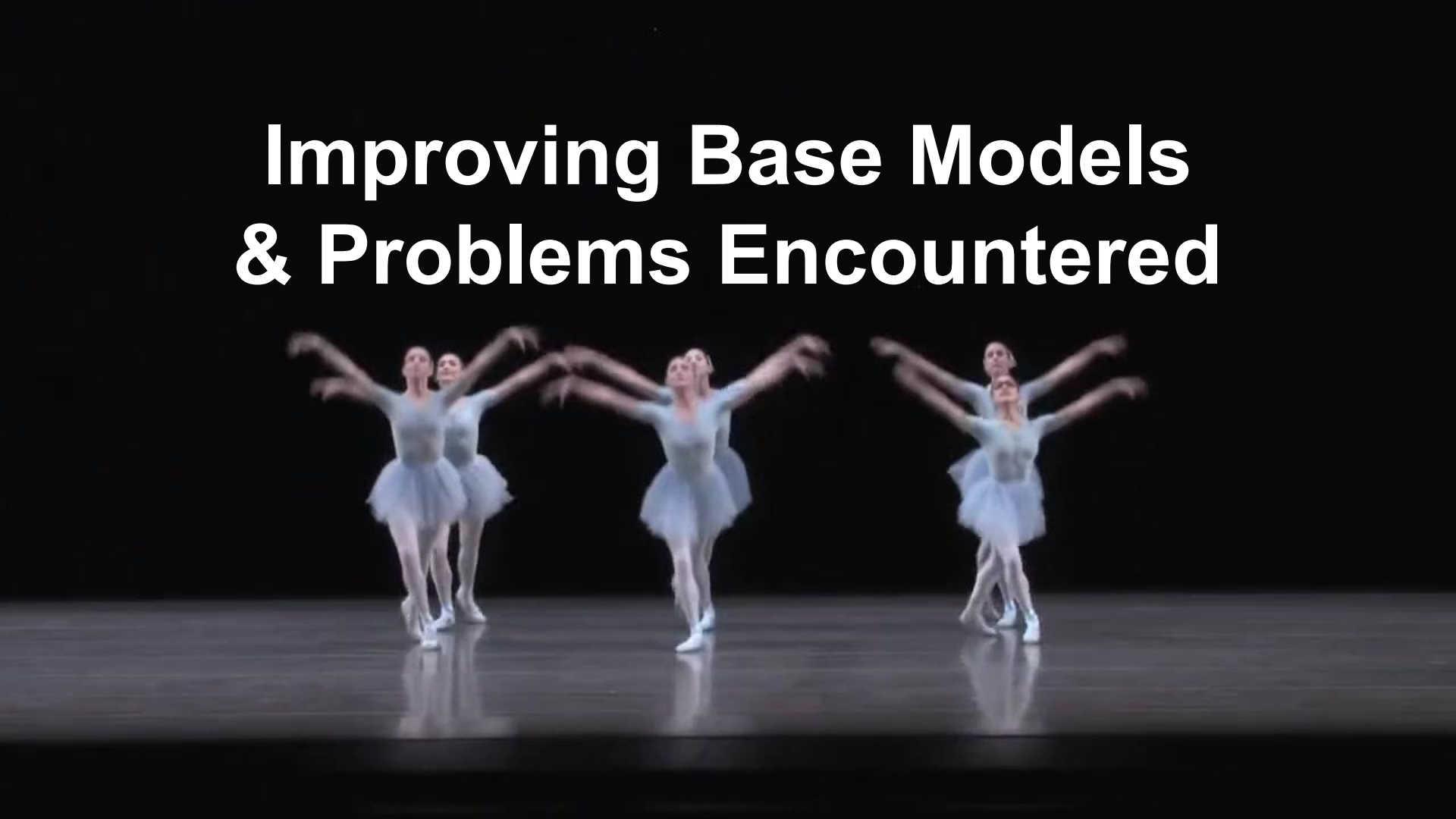
Properly Split Train & Test Data



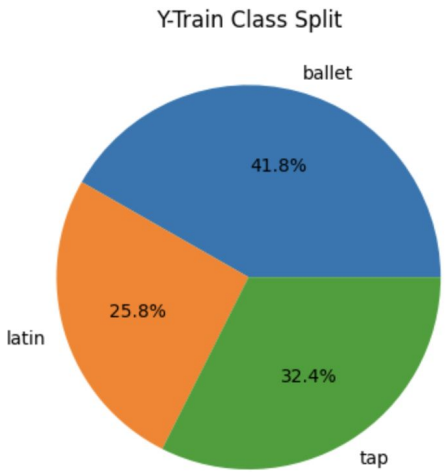


Model	% Accuracy (train/test split on **images**)	% Accuracy (train/test split on **videos**)
Random Forest	96%	58%
Neural Net	84%	50%
k-Nearest Neighbor	96%	45%

Improving Base Models & Problems Encountered



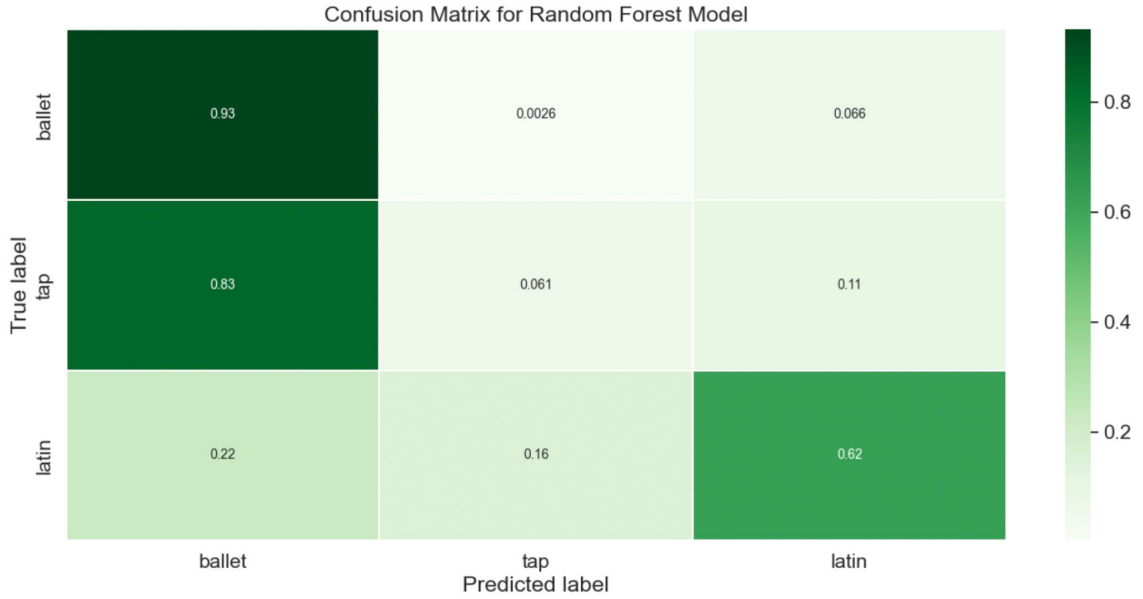
Random forest: 3 Class Model



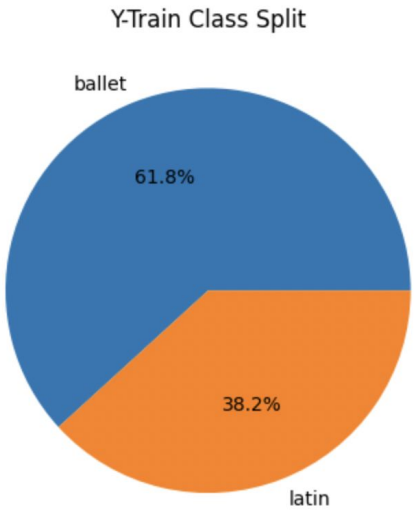
Baseline Accuracy: 42%

RF Train Accuracy: 63%

RF Test Accuracy: 58%



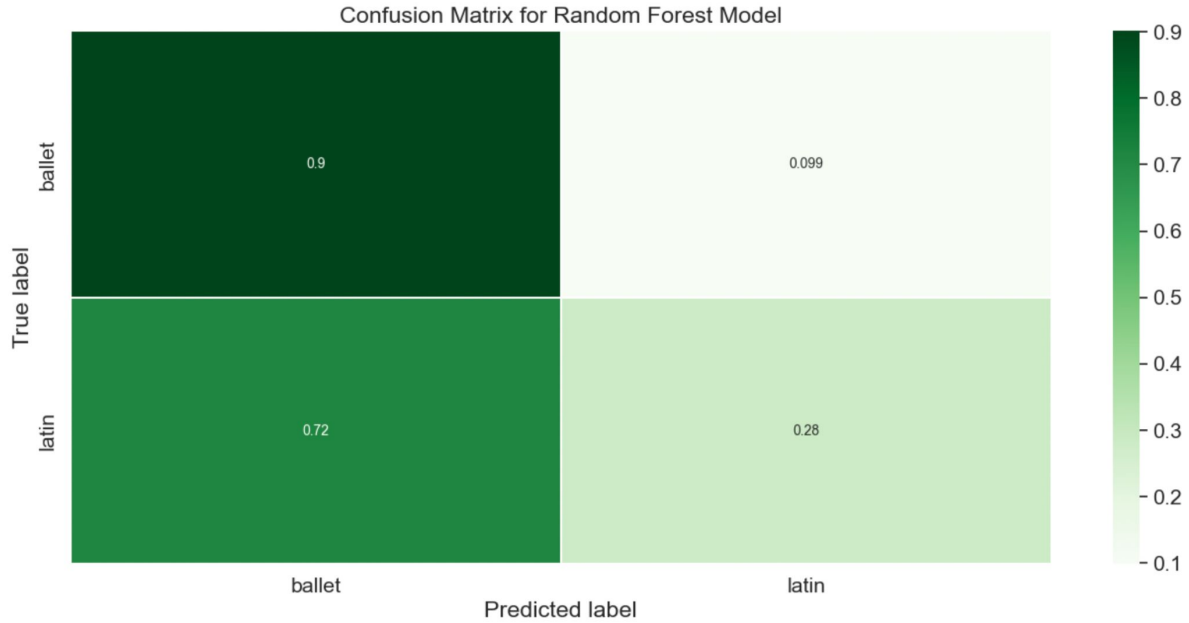
Random forest: 2 Class Model



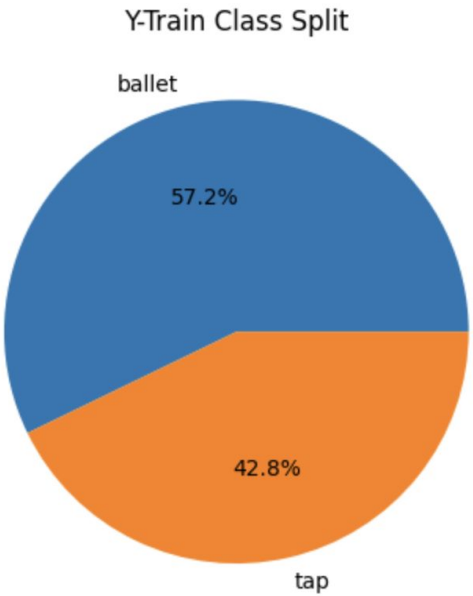
Baseline Accuracy: 62%

RF Train Accuracy: 89%

RF Test Accuracy: 65%



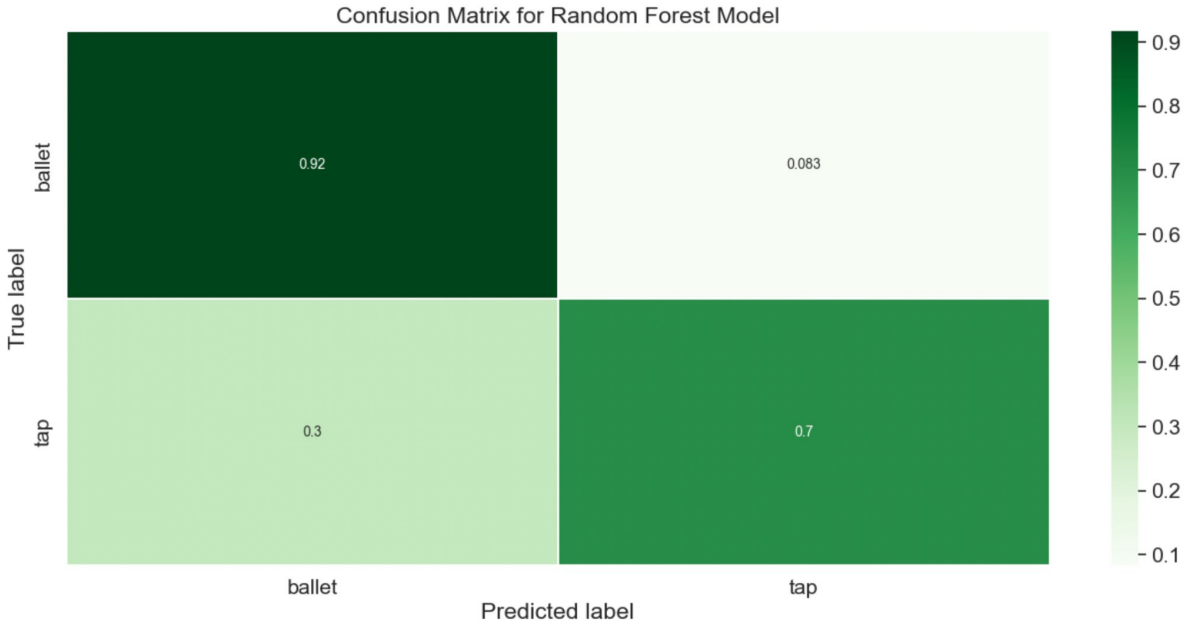
Random Forest: Another 2-Class Model (Best)



Baseline Accuracy: 57%

RF Train Accuracy: 86%

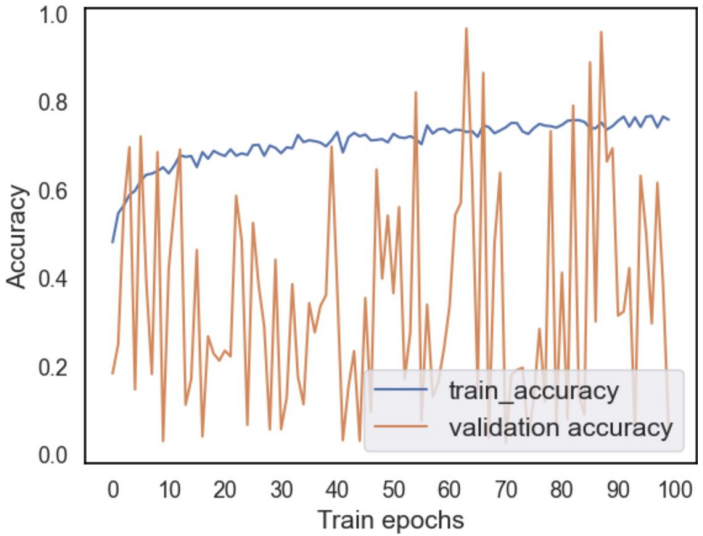
RF Test Accuracy: 79%



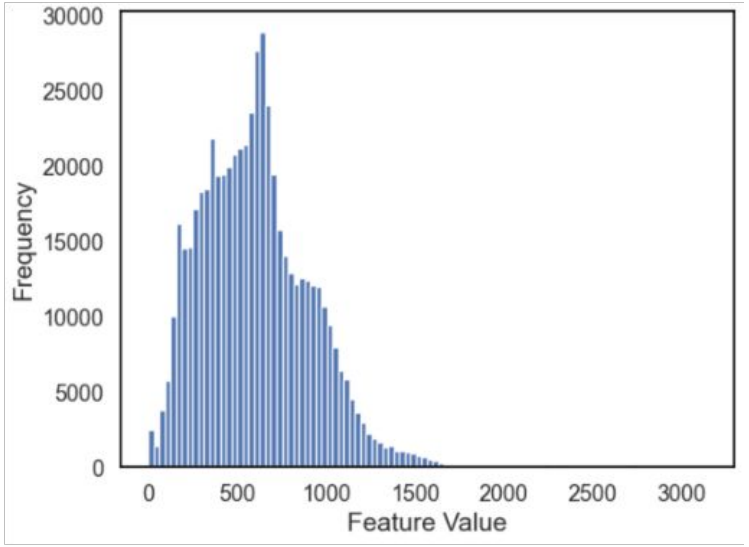
Neural Net - 3 Class Model

Training and Validation Accuracy of NN

- 3 classes (Ballet, Tap, Latin)
- Train/test split separated by video



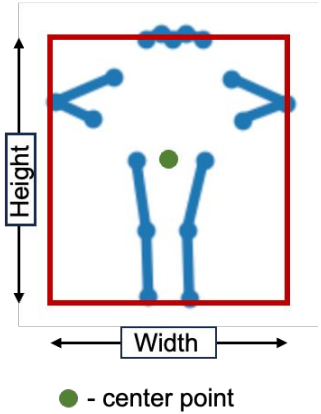
Frequency of (x,y) feature size

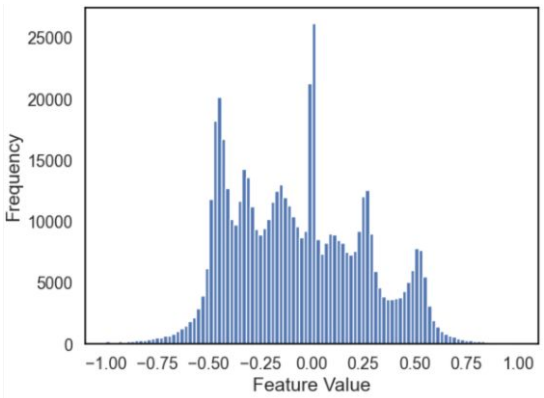


- Extremely noisy validation accuracy
- Likely caused by large range of feature values
  - Depends strongly camera zoom level

Normalization & Scaling

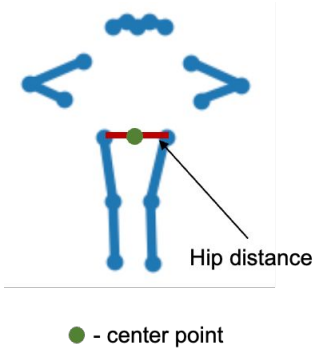
Bounding Box Method

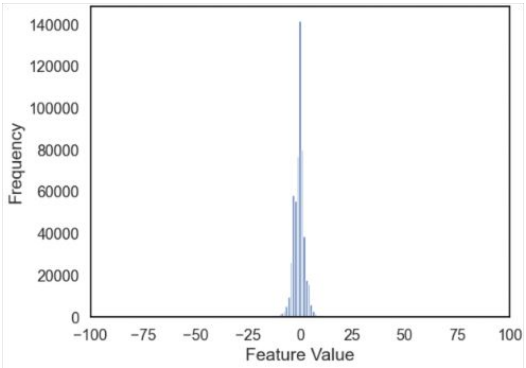




- Center based on center of hips
- Scale x values by width, y values by height

Hip Distance Method

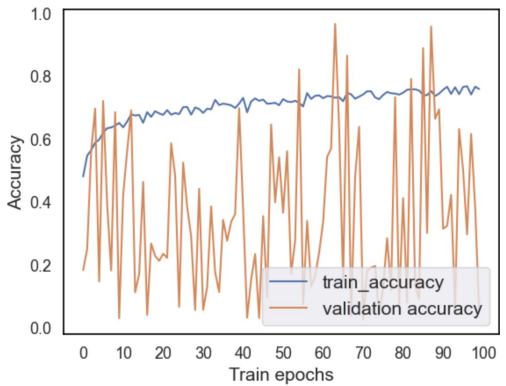




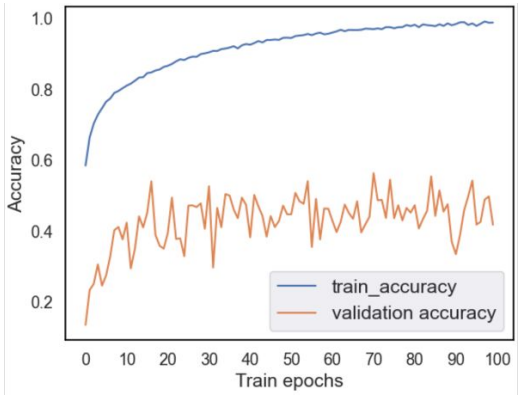
- Center based on center of hips
- Scale x and y based on hip width

Normalization Impact on Training

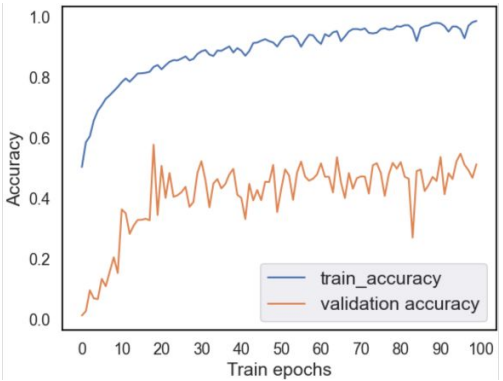
Non-normalized features



Bounding Box



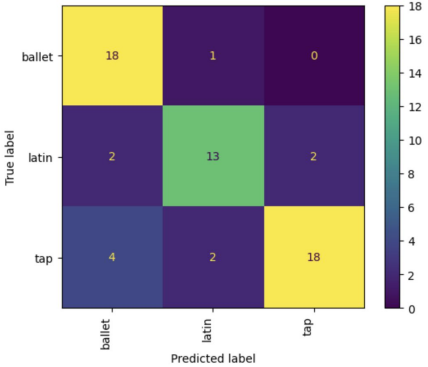
Hip Distance



Reduced noise, but did not improve accuracy

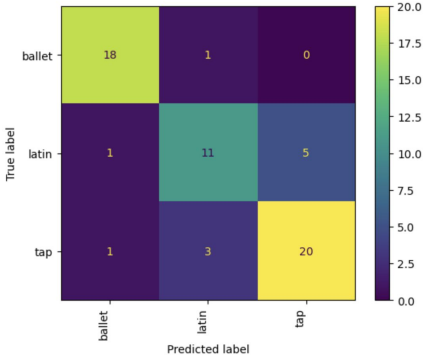

Expanding the Dataset

Raw images



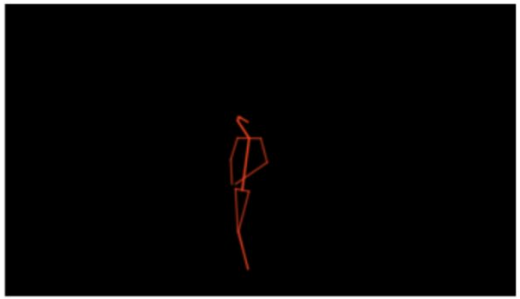
Random Forest
(n_estimators=2, n_jobs=-1, random_state=7, max_depth=8)
Accuracy for Training: **0.8861111111111111**
Accuracy for Validation: **0.8166666666666667**

Random Forest
(n_estimators=180, n_jobs=-1, random_state=7, max_depth=8)
Accuracy for Training: **1.0**
Accuracy for Validation: **0.8166666666666667**

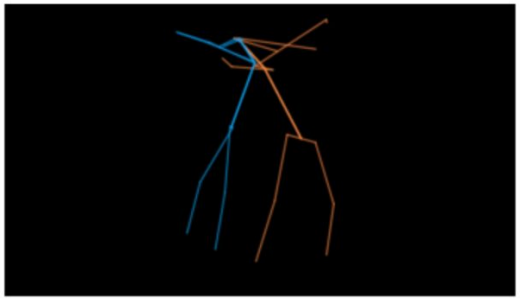


Skeletal images

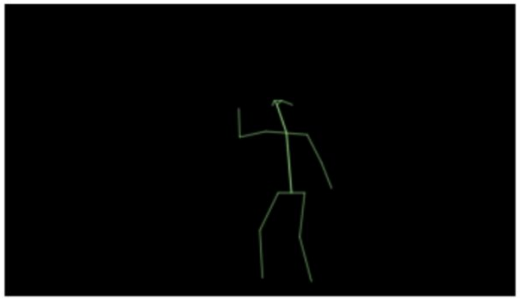
ballet

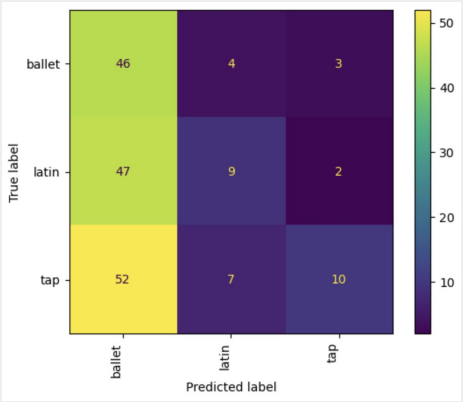


latin



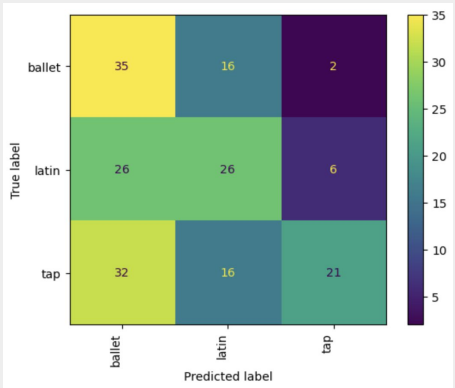
tap



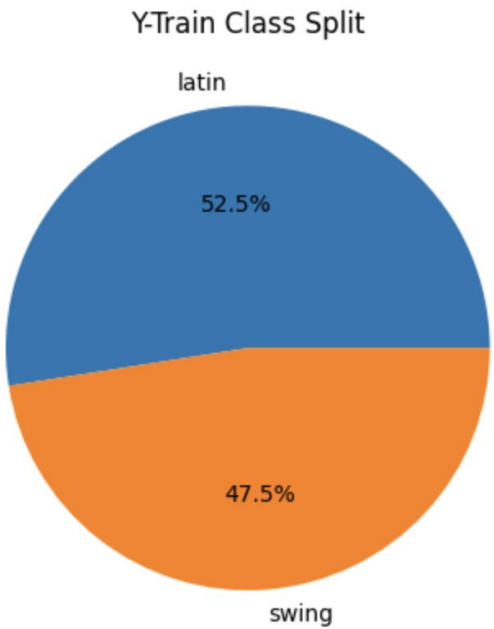


Accuracy for Training:
0.5838888888888889
Accuracy for Validation:
0.35833333333333334

Accuracy for Training:
0.7933333333333333
Accuracy for Validation:
0.5183333333333333



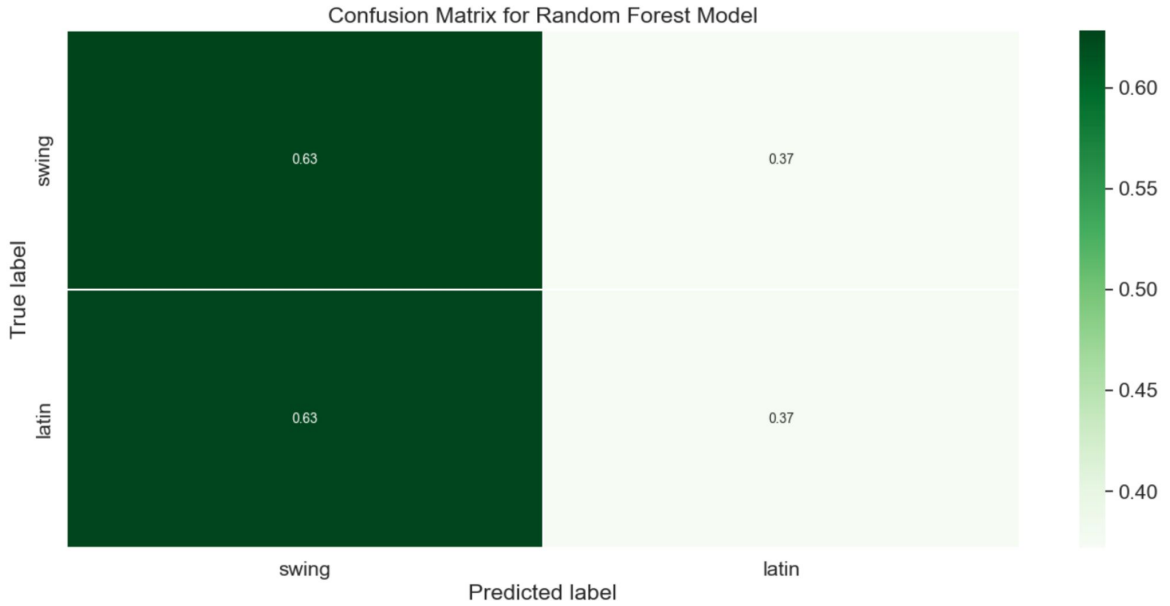
Random Forest: 2-Person 2-Class Model



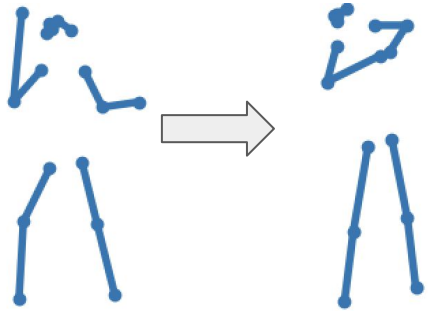
Baseline Accuracy: 53%

RF Train Accuracy: 97%

RF Test Accuracy: 50%



Temporal Classification



Methodology:

- Expand model scope to include temporal elements
- **Dance Movement**: the difference between body party coordinates in a frame and the corresponding frame 10 frames ahead
- Established pairings of image frames and their 10-frame-ahead counterparts
- New feature matrix has 68 columns (34 for current frame + 34 for temporal association)

	nose_x	nose_y	left_eye_x	left_eye_y	right_eye_x	right_eye_y	left_ear_x	left_ear_y	right_ear_x	right_ear_y	...	right_hip_x_tmp	right_hip_y_tmp	left_knee_x_tmp
0	969.196716	620.896118	980.181335	601.907410	968.198120	603.906189	1021.124084	603.906189	968.198120	605.905029	...	84.084900	-103.438904	-145.534485
1	993.960449	569.382507	1008.953491	559.395630	994.959961	557.398315	1047.935425	576.373291	992.960938	569.382507	...	17.509827	-83.663391	-10.505859
2	1020.045349	536.810242	1035.035156	522.820862	1016.048096	522.820862	1072.009766	535.811035	1012.050842	537.809509	...	4.623291	-86.960205	-6.584106
3	1029.675903	520.397827	1043.664551	504.398682	1023.680786	509.398407	1079.635254	521.397766	1017.685669	525.397522	...	8.158691	-104.476257	-15.798340
4	1035.588867	515.905334	1049.573730	495.930664	1037.586670	494.931946	1082.537964	516.904053	1025.599731	527.890076	...	-3.593140	-46.221069	-16.608887
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
13558	949.987427	289.156189	952.985657	280.163513	941.992188	280.163513	909.011841	280.163513	911.010620	280.163513	...	-92.922729	-13.120880	-175.898682
13559	928.200562	283.913574	933.194885	275.920624	920.209717	274.921509	936.191406	281.915344	896.237061	274.921509	...	-87.490906	-7.952362	-164.431091
13560	897.908386	272.676544	906.870850	265.683868	890.937622	263.685974	876.996033	266.682831	876.996033	267.681793	...	-96.623901	2.413818	-146.400452
13561	849.859131	271.653595	860.850525	262.666687	851.857544	261.668121	894.823975	272.652130	897.821655	272.652130	...	-125.947144	-18.216187	-141.015259
13562	829.694702	269.840942	836.676819	249.852402	832.687012	245.854691	859.618103	262.844940	885.551697	272.839233	...	-127.577026	-7.278473	-79.752136

13563 rows × 68 columns

Temporal Classification

Model Performance

- Random Forest accuracy: ~65%
- Neural Network accuracy: ~66%

CONCLUSION

Reflections and Future Directions

- Great team collaboration and contribution
- Achieved approximately 80% accuracy with a slightly tuned Random Forest model!
- Our Neural Net model accuracy was boosted by over 15% for our temporal dataset!
  - Neural Net Model on single person video 3 class classification: ~50% accuracy
  - Neural Net Model on temporal data 3 class classification: ~66% accuracy
- Originally aimed for a dance move generator, but hey things don't always turn out the way we expect → adaptation is key
- Future goals:
  - Utilize more of the dataset
  - Classify videos with multiple individuals
  - ULTIMATELY: create choreography-generating model :)