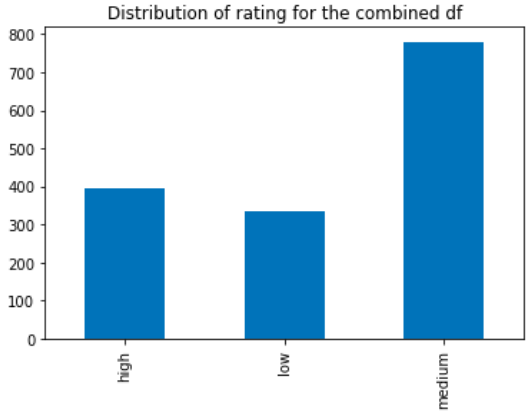
RANDOM FOREST (BAGGING) RESULTS

Notes:

- When predicting properties, the order of the labels for the confusion matrix are
  1. Smoothness
  2. Thickness
  3. Warmth
  4. Flexibility
  5. Softness
- When predicting rating, the order of the labels for the confusion matrix are
  1. low
  2. medium
  3. high
- Generated random seed for random forest is 183
- When predicting properties, the macro F1 score was used as the data was balanced (same number of observations for each property)
- When predicting ratings, the weighted F1 score was used as the data was unbalanced (refer figure below)



(1) Estimating the PROPERTY based on available data:

- Chance classification accuracy: 20% (1/5)

Method 1a: Leave One Participant Out CV using all 180 features

Confusion matrix	Average Micro F1 score	Average Macro F1 score	Overall accuracy
[[132 40 26 43 11] [36 107 16 55 38] [62 26 54 66 44] [33 41 18 122 38] [56 62 14 66 54]]	0.37	0.32	37.22%

Participant left out	Number of trees in rf	Confusion matrix	Micro F1 score	Macro F1 score	Classification accuracy
1	500	[[9 6 3 0 0] [1 9 6 0 2] [6 0 12 0 0] [2 11 0 0 5] [2 16 0 0 0]]	0.33	0.28	33.33%
2	1000	[[11 3 0 3 1] [0 18 0 0 0] [0 0 18 0 0] [2 0 0 16 0] [4 7 0 6 1]]	0.71	0.65	71.11%
3	500	[[11 0 4 3 0] [0 10 0 7 1] [2 2 0 13 1] [1 0 0 17 0] [0 8 0 7 3]]	0.46	0.4	45.56%
4	1000	[[9 0 4 2 3] [0 9 0 6 3] [1 0 0 8 9] [0 0 0 15 3] [0 0 0 10 8]]	0.46	0.44	45.56%
5	1000	[[12 0 4 2 0] [3 3 0 0 12] [13 0 2 0 3] [0 0 1 17 0] [0 0 0 0 18]]	0.58	0.52	57.78%
6	1000	[[11 0 1 2 4] [0 8 0 3 7] [0 7 3 0 8] [3 0 0 7 8] [2 0 2 2 12]]	0.46	0.45	45.56%
7	1500	[[10 8 0 0 0] [0 16 0 0 2] [8 7 3 0 0] [2 13 0 1 2] [4 12 0 2 0]]	0.33	0.26	33.33%
8	1500	[[3 0 0 14 1] [0 0 0 15 3] [5 0 0 13 0] [3 1 0 14 0] [0 0 0 16 2]]	0.21	0.14	21.11%

19	1500	[[1 9 2 6 0] [0 6 0 12 0] [0 2 0 14 2] [0 2 0 16 0] [1 6 0 10 1]]	0.27	0.18	26.67%
21	1500	[[17 0 1 0 0] [17 0 1 0 0] [16 0 0 2 0] [17 1 0 0 0] [18 0 0 0 0]]	0.19	0.07	18.89%
22	500	[[8 7 2 0 1] [4 11 3 0 0] [2 6 7 0 3] [2 7 5 0 4] [3 5 3 3 4]]	0.33	0.29	33.33%
23	500	[[13 2 1 1 1] [8 2 0 5 3] [8 0 1 6 3] [0 3 1 8 6] [8 5 0 3 2]]	0.29	0.24	28.89%
24	1500	[[15 3 0 0 0] [3 9 1 0 5] [1 2 0 0 15] [1 3 3 1 10] [13 1 1 0 3]]	0.31	0.26	31.11%
25	1500	[[2 2 4 10 0] [0 6 5 7 0] [0 0 8 10 0] [0 0 8 10 0] [1 2 8 7 0]]	0.29	0.25	28.89%

Method 1b: Leave One Participant Out CV using only the emg features (48 features)

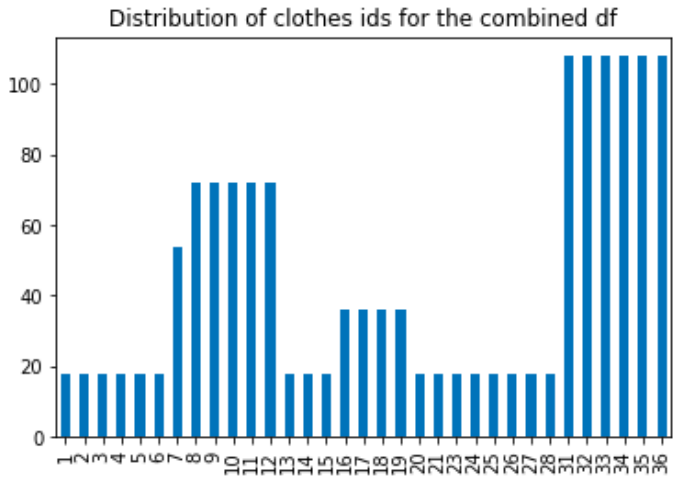
Confusion matrix	**Average Micro F1 score**	**Average Macro F1 score**	**Overall accuracy**
[[129 24 53 28 18] [65 62 35 65 25] [93 28 57 36 38] [47 28 25 120 32] [55 42 39 74 42]]	0.33	0.28	32.54%

Participant left out	Number of trees in rf	Confusion matrix	Micro F1 score	Macro F1 score	Classification accuracy
1	1500	[[18 0 0 0 0] [6 3 5 0 4] [18 0 0 0 0] [5 3 2 2 6] [7 4 0 0 7]]	0.33	0.26	33.33%
2	500	[[8 4 3 3 0] [5 5 2 4 2] [5 5 8 0 0] [4 3 1 4 6] [2 6 1 9 0]]	0.28	0.26	27.78%
3	500	[[6 0 12 0 0] [0 11 3 4 0] [7 1 5 4 1] [0 9 0 9 0] [0 12 1 5 0]]	0.34	0.3	34.44%
4	1500	[[7 0 3 2 6] [1 4 0 7 6] [2 1 4 1 10] [1 0 0 11 6] [0 1 0 4 13]]	0.43	0.42	43.33%
5	1000	[[0 0 15 3 0] [0 3 3 12 0] [2 0 12 1 3] [0 1 0 14 3] [0 0 4 4 10]]	0.43	0.37	43.33%
6	1000	[[12 1 2 3 0] [6 4 0 7 1] [9 2 3 2 2] [3 2 0 9 4] [3 0 1 9 5]]	0.37	0.35	36.67%
7	1000	[[18 0 0 0 0] [10 3 5 0 0] [14 0 3 0 1] [7 1 5 3 2] [1 2 10 5 0]]	0.3	0.23	30.0%
8	500	[[3 3 1 1 10] [0 3 1 6 8] [3 7 1 1 6] [0 4 0 14 0] [0 1 2 12 3]]	0.27	0.24	26.67%
19	500	[[8 8 1 0 1] [1 9 1 6 1]	0.4	0.35	40.0%

		[2 4 1 6 5] [0 1 0 16 1] [5 8 0 3 2]]			
21	500	[[14 0 4 0 0] [16 0 2 0 0] [17 0 0 0 1] [18 0 0 0 0] [12 0 6 0 0]]	0.16	0.06	15.56%
22	500	[[10 4 4 0 0] [4 4 8 0 2] [3 4 10 0 1] [1 3 12 0 2] [5 3 9 1 0]]	0.27	0.21	26.67%
23	1500	[[10 3 0 4 1] [3 7 2 6 0] [5 2 1 10 0] [1 0 1 16 0] [6 2 0 8 2]]	0.4	0.34	40.0%
24	1500	[[13 0 5 0 0] [11 2 3 1 1] [5 1 4 0 8] [7 1 2 6 2] [12 1 5 0 0]]	0.28	0.25	27.78%
25	1500	[[2 1 3 12 0] [2 4 0 12 0] [1 1 5 11 0] [0 0 2 16 0] [2 2 0 14 0]]	0.3	0.24	30.0%

Method 2: Leave One Cloth Out CV

Made sense to do LOCOCV for the sock data as every participant touched all socks. However, the clothes that the participants in Lili's experiment touched were different. So thought LOCOCV might not make sense.



(2) Estimating the RATING based on data:

As the data is imbalanced used the weighted F1 score instead of the macro F1 score

- Chance classification accuracy: 33.33% (1/3)

Method 1a: Leave One Participant Out CV using all 180 features

Property	Confusion matrix	Average micro F1 score	Average weighted F1 score	Average classification accuracy
Smoothness	[[0 55 2] [11 120 10] [2 49 3]]	0.49	0.39	48.81%
Thickness	[[6 65 1] [23 122 2] [3 30 0]]	0.51	0.44	50.79%
Warmth	[[0 32 7] [0 108 30] [0 67 8]]	0.46	0.38	46.03%
Flexibility	[[26 31 15] [21 41 46] [8 42 22]]	0.35	0.34	35.32%

Softness	[[5 36 19] [15 69 24] [12 29 43]]	0.46	0.38	46.43%

Method 1b: Leave One Participant Out CV using only the emg features (48 features)

Property	**Confusion matrix**	**Average micro F1 score**	**Average weighted F1 score**	**Average classification accuracy**
Smoothness	[[4 52 1] [9 129 3] [3 51 0]]	0.53	0.41	52.78%
Thickness	[[6 66 0] [21 126 0] [3 30 0]]	0.52	0.42	52.38%
Warmth	[[0 36 3] [0 119 19] [0 65 10]]	0.51	0.4	51.19%
Flexibility	[[14 48 10] [28 63 17] [17 34 21]]	0.39	0.36	38.89%
Softness	[[3 32 25] [21 65 22] [16 54 14]]	0.33	0.28	32.54%

Method 2: Leave One Sock Out CV

Didn't do this for the combined dataset for the reason stated above