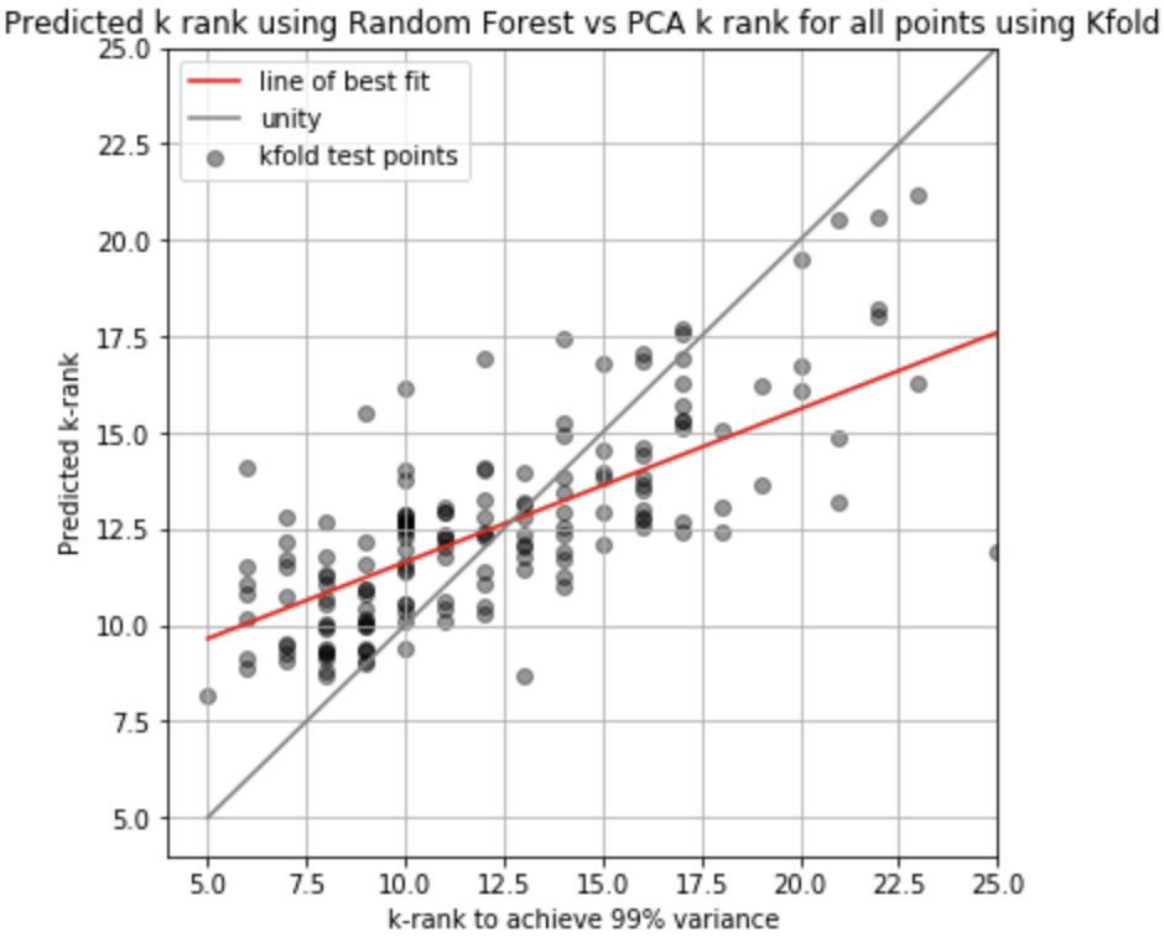
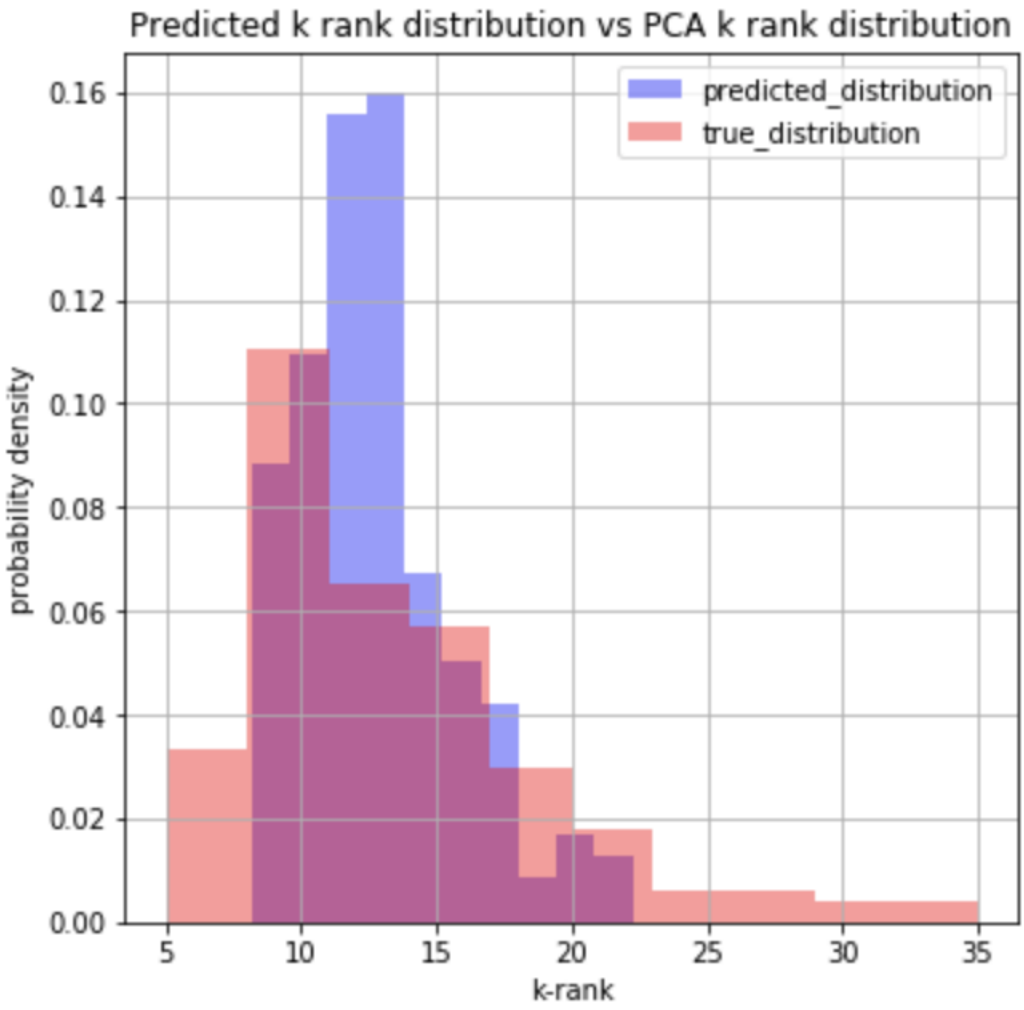
Feature importance for RandomForest (information gain for each feature, therefore sum to 1):

```
{'binned_complexity': 0.12,
 'horizontal_complexity': 0.08,
 'length_soma': 0.03,
 'avg_length_dendrite': 0.03,
 'avg_length_axon': 0.01,
 'total_length_dendrite': 0.03,
 'total_length_axon': 0.01,
 'length_dendrite_var': 0.03,
 'length_axon_var': 0.01,
 'count_axon': 0.01,
 'count_dendrite': 0.0,
 'avg_diameter_dendrite': 0.01,
 'avg_diameter_axon': 0.03,
 'var_diameter_dendrite': 0.03,
 'var_diameter_axon': 0.0,
 'm_type': 0.23,
 'e_type': 0.34,
 'level': 0.0}
```



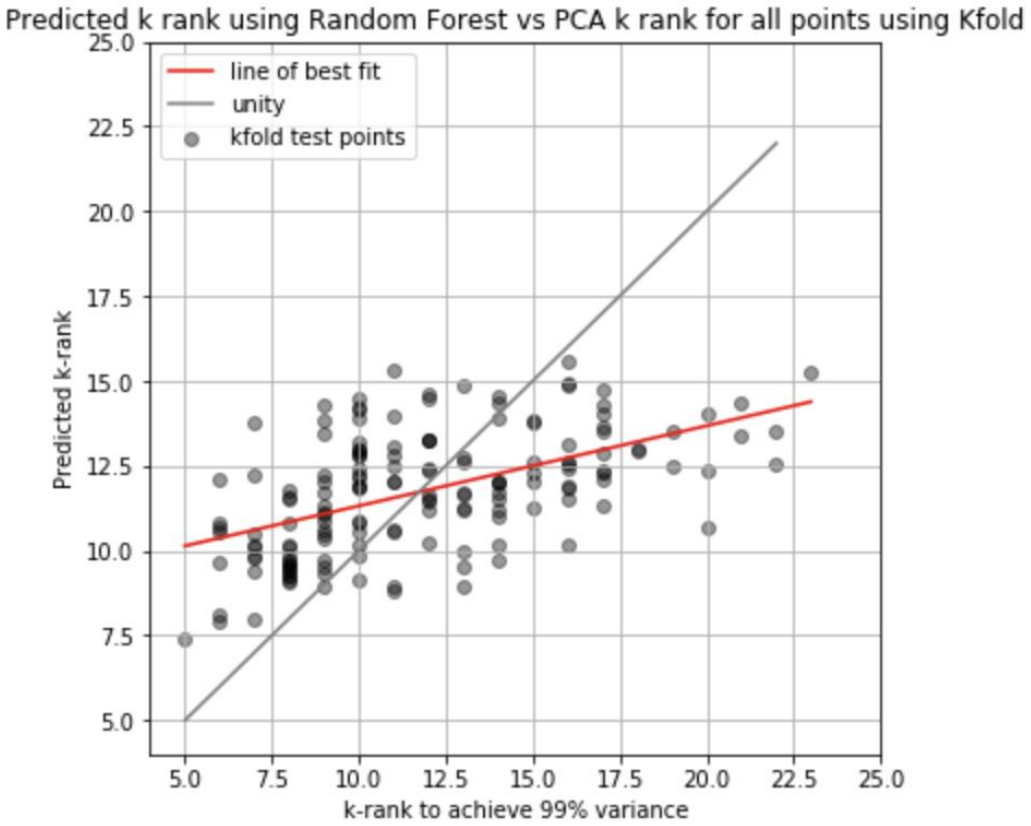
R2 = 0.52


Predicted k rank distribution vs PCA k rank distribution
predicted_distribution
true_distribution
probability density
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
5
10
15
20
25
30
35
k-rank

Removing E-type, M-type and level (i.e. categorized data):

Feature importance

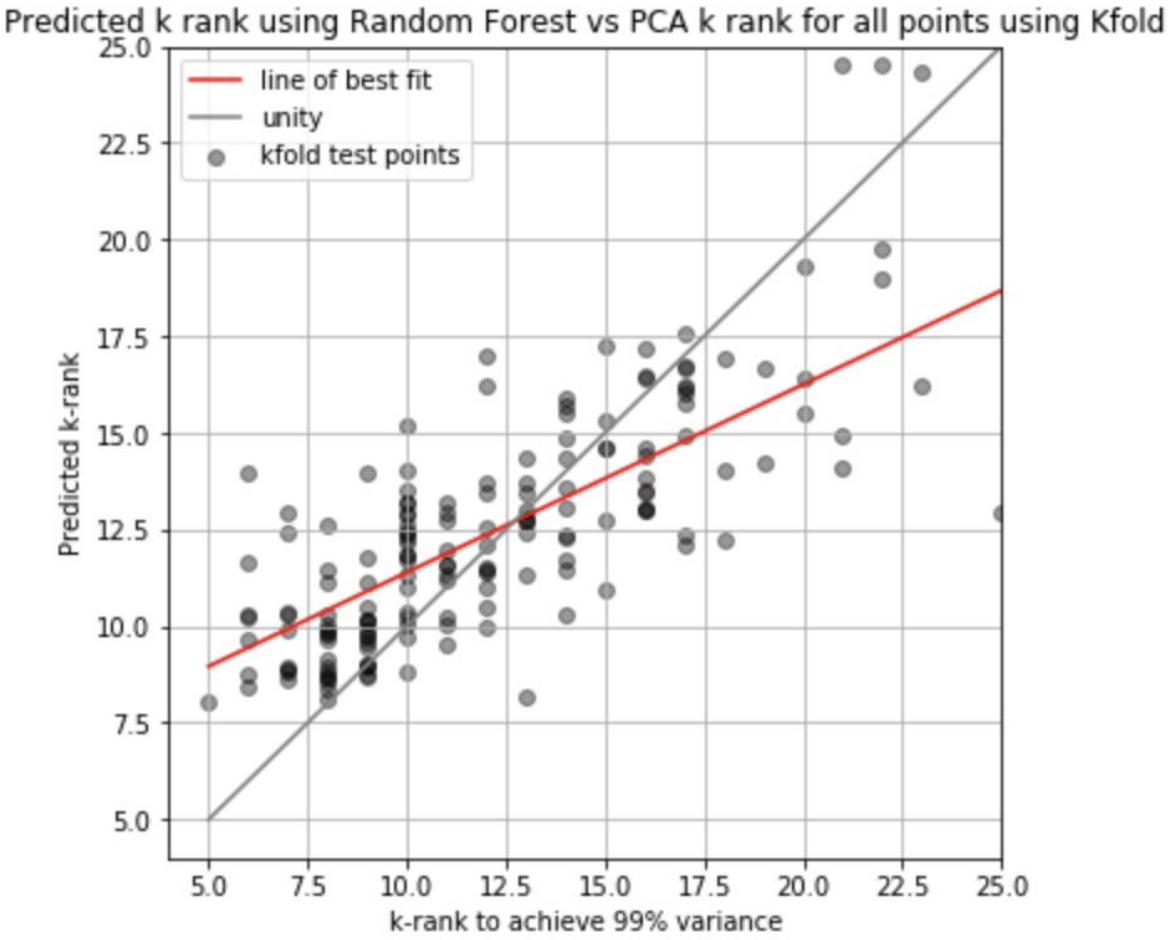
{'binned_complexity': 0.15,
 'horizontal_complexity': 0.04,
 'length_soma': 0.16,
 'avg_length_dendrite': 0.03,
 'avg_length_axon': 0.04,
 'total_length_dendrite': 0.06,
 'total_length_axon': 0.04,
 'length_dendrite_var': 0.03,
 'length_axon_var': 0.04,
 'count_axon': 0.02,
 'count_dendrite': 0.03,
 'avg_diameter_dendrite': 0.08,
 'avg_diameter_axon': 0.08,
 'var_diameter_dendrite': 0.14,
 'var_diameter_axon': 0.06}


Predicted k rank using Random Forest vs PCA k rank for all points using Kfold
line of best fit
unity
kfold test points
Predicted k-rank
k-rank to achieve 99% variance
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0

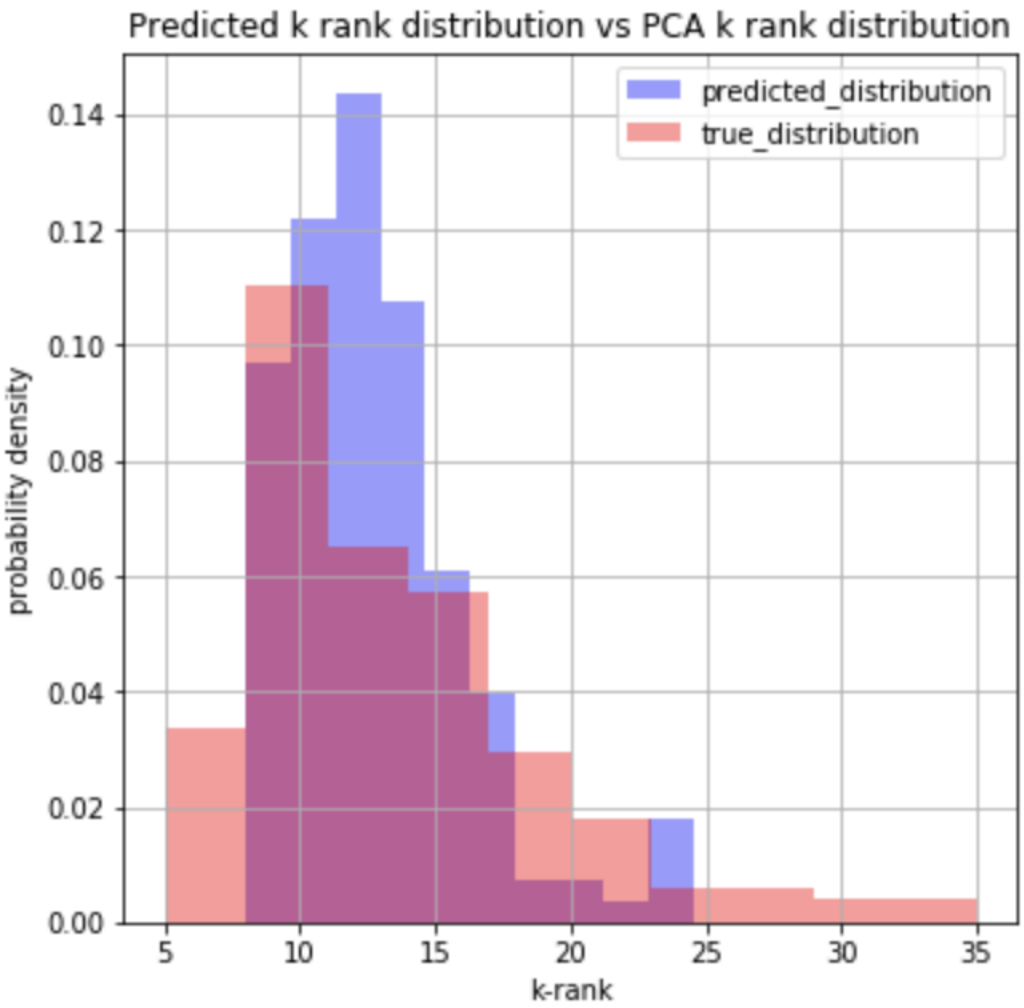
R2 = 0.26

Choosing only feature provided by Lasso

```
{'binned_complexity': 0.13,
 'length_soma': 0.03,
 'avg_length_dendrite': 0.04,
 'avg_length_axon': 0.01,
 'total_length_dendrite': 0.05,
 'length_axon_var': 0.01,
 'avg_diameter_axon': 0.03,
 'var_diameter_axon': 0.0,
 'm_type': 0.26,
 'e_type': 0.43,
 'levels': 0.0}
```



R2= 0.59


Predicted k rank distribution vs PCA k rank distribution
predicted_distribution
true_distribution
probability density
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
5
10
15
20
25
30
35
k-rank