

## CH Activation - Best Model Feature Analysis

### Published Model

#### 3rd Place

formula

class ~ fr\_NH0 + fr\_aldehyde + fr\_amide + fr\_pyridine

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy    McFadden\_R2

85.42%    0.762

Model Coefficients

	(Interc ept)	fr_NH0 1	fr_NH 02	fr_NH0 5	fr_aldehy de1	fr_amid e1	fr_amid e2	fr_pyridi ne1	fr_pyridi ne2
2	77.146 276	7.6137 79	- 76.42 688	- 11.946 103	0.84884 33	- 56.072 04	- 20.527 82	- 86.2502 4	78.0081 71
3	8.3193 67	- 52.025 492	- 57.35 079	- 3.9100 05	- 185.047 3453	103.38 214	111.78 747	- 162.834 43	- 3.61883 8
4	95.677 264	- 51.465 887	- 95.25 421	24.534 540	- 196.787 2443	14.771 46	- 72.367 90	- 155.401 07	- 33.2751 26

## Training Set

### Confusion Matrix

A	9 (90%)	1 (10%)	0 (0%)	0 (0%)	10 (20.8%)
B	2 (14.3%)	12 (85.7%)	0 (0%)	0 (0%)	14 (29.2%)
C	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
D	0 (0%)	1 (7.1%)	3 (21.4%)	10 (71.4%)	14 (29.2%)
total	11 (81.8%)	14 (85.7%)	13 (76.9%)	10 (100%)	48 (85.4%)
	A	B	C	D	total

Proportion

0  
25  
50  
75  
100

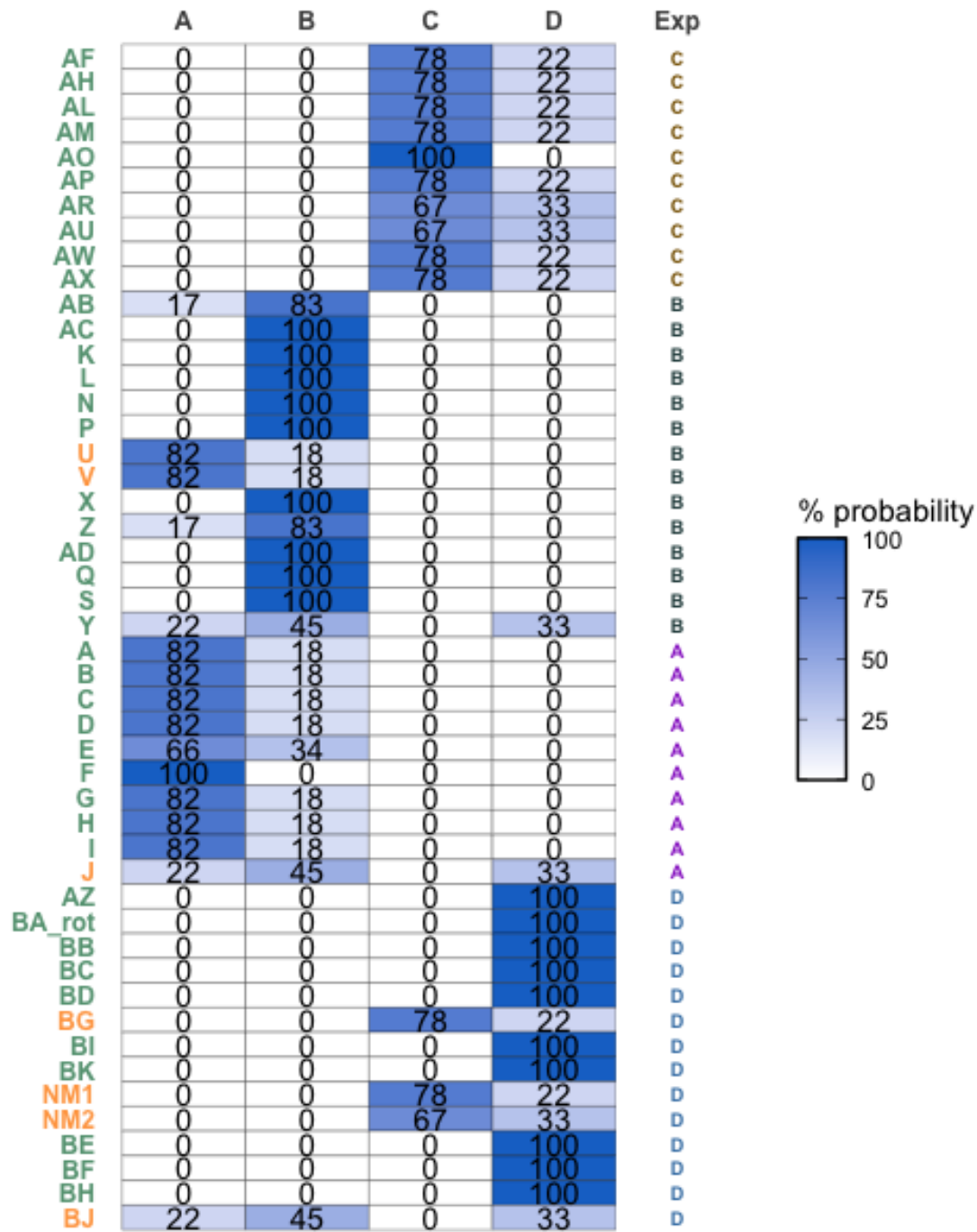
Legend

Accuracy  
False  
Precision  
Size  
True

3. 3rd place - 4 features

## Training Set

### Probability Heatmap




3. 3rd place - 4 features

## Test Set

### Confusion Matrix


A	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
B	0 (0%)	3 (100%)	0 (0%)	0 (0%)	3 (50%)
C	0 (0%)	0 (0%)	3 (100%)	0 (0%)	3 (50%)
D	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
total	0 (0%)	3 (100%)	3 (100%)	0 (0%)	6 (100%)
	A	B	C	D	total

Proportion



0  
25  
50  
75  
100

Legend



Accuracy  
False  
Precision  
Size  
True

3. 3rd place - 4 features

### Proportion

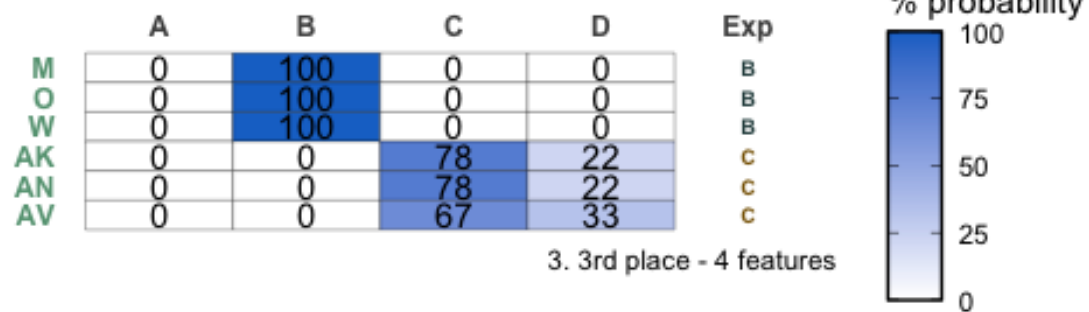


### Legend



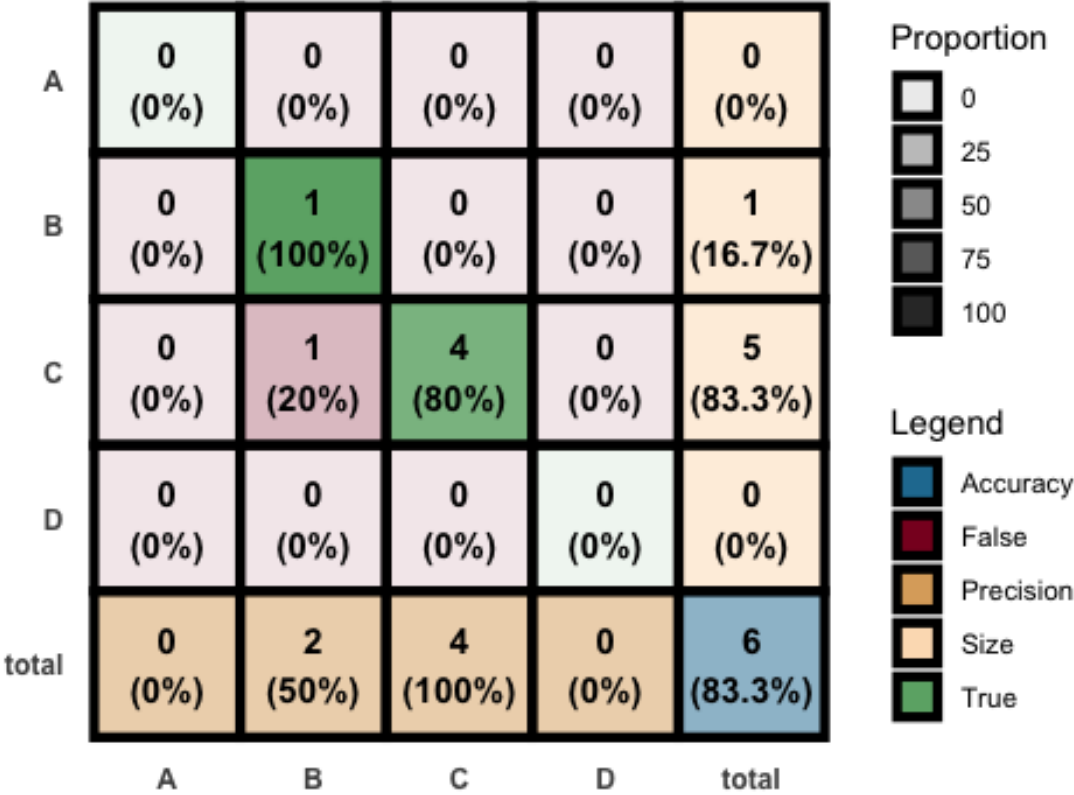
## Test Set

### Probability Heatmap



External Validation

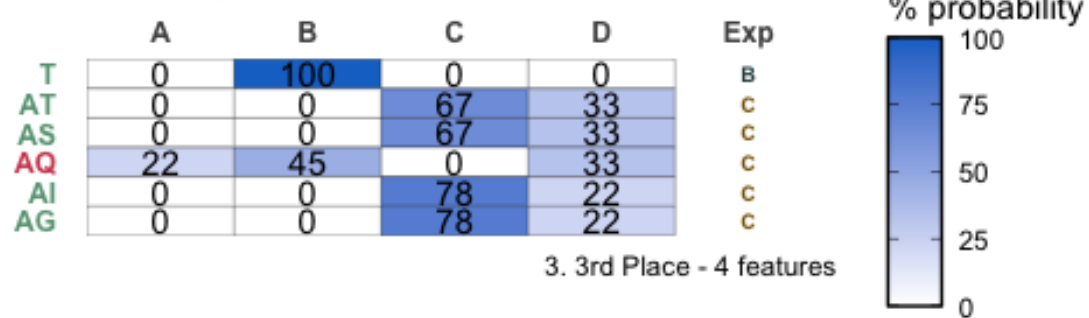
Confusion Matrix



3. 3rd Place - 4 features

## External Validation

### Probability Heatmap



### External Validation

	Prediction	1	2	3	4
p1	4	0	0	0	100
p2	1	100	0	0	0
p3	2	22	45	0	33
p4	2	0	100	0	0
p5	2	22	45	0	33
p6	2	22	45	0	33
p7	2	22	45	0	33
p8	2	22	45	0	33
p9	2	22	45	0	33
p10	2	0	100	0	0

### Model Analysis - Variable Impact

#### 1. fr\_NH0 Removed

formula

---

class ~ fr\_aldehyde + fr\_amide + fr\_pyridine

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
----------	-------------

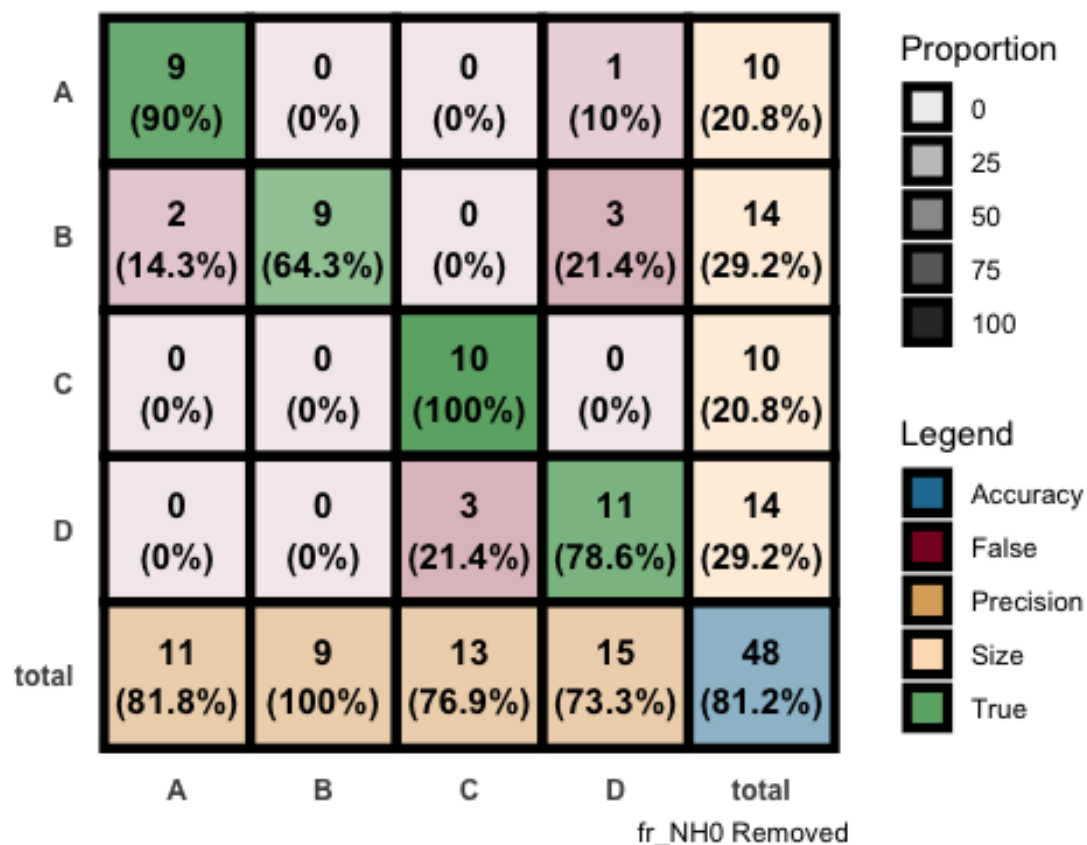
81.25%	0.639
--------	-------

Model Coefficients

	(Intercept)	fr_aldehyde1	fr_amide1	fr_amide2	fr_pyridine1	fr_pyridine2
2	1.761409	1.263614	-0.7059781	-14.79323	-3.373458	23.695461
3	-16.420132	-56.180766	51.2368906	47.97627	-59.021896	-1.446856
4	2.931482	-65.434124	30.7865976	-21.31971	-46.467270	-15.186553

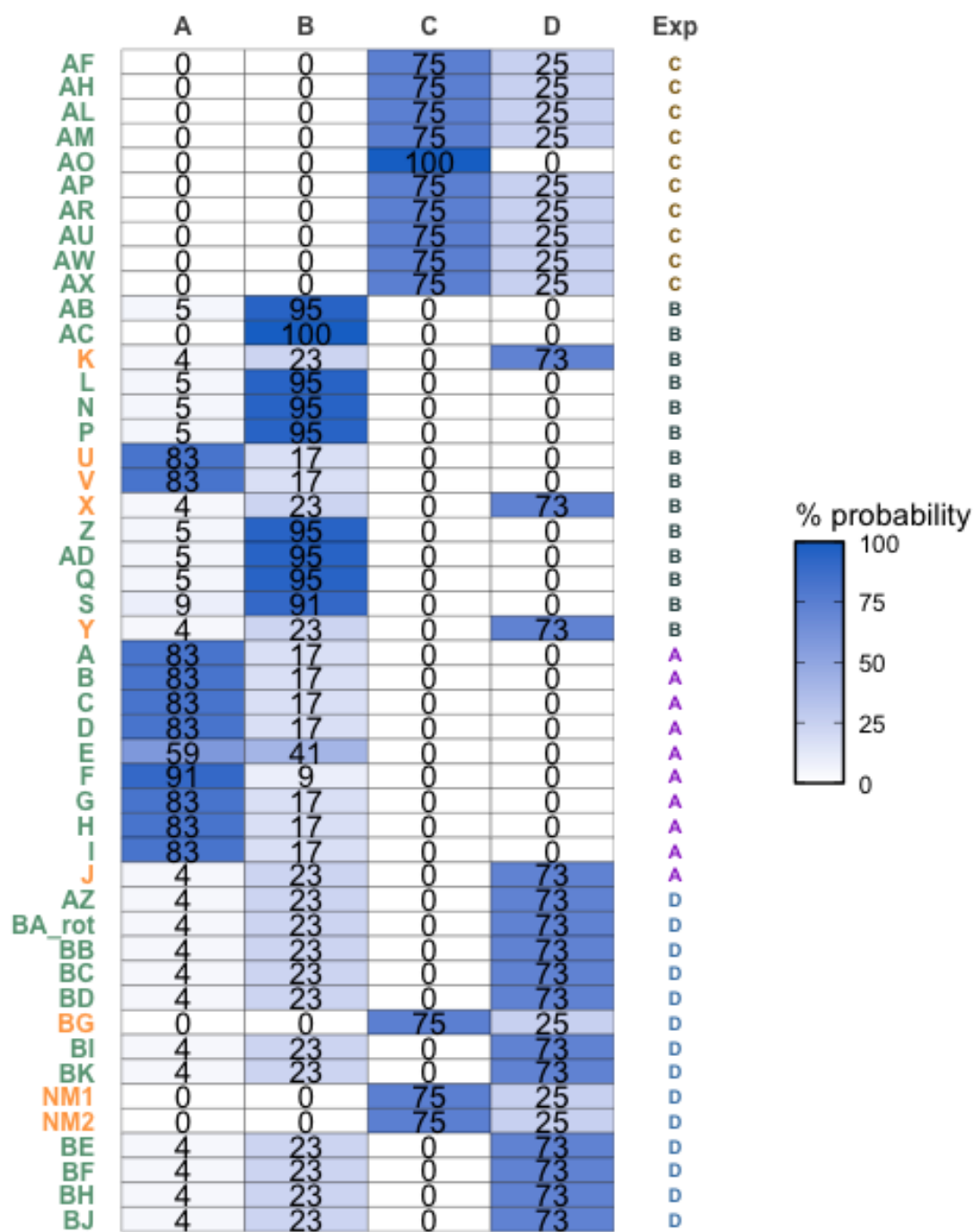


# Training Set Confusion Matrix



# Training Set

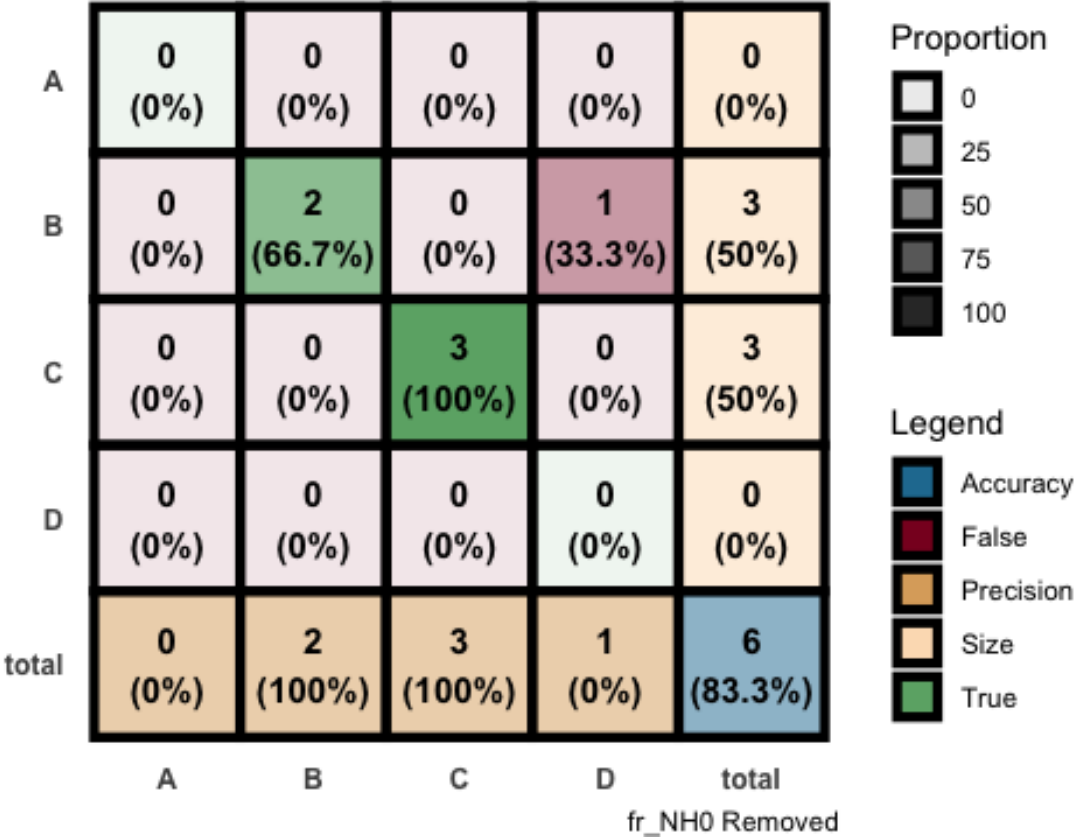
## Probability Heatmap



fr\_NH0 Removed

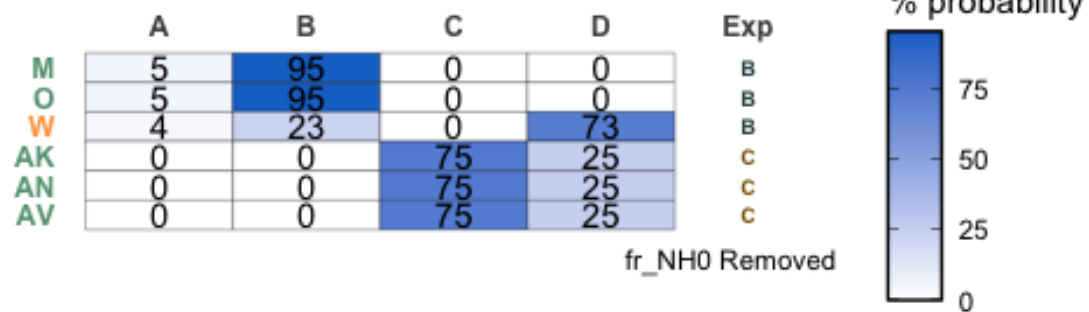
Test Set

Confusion Matrix



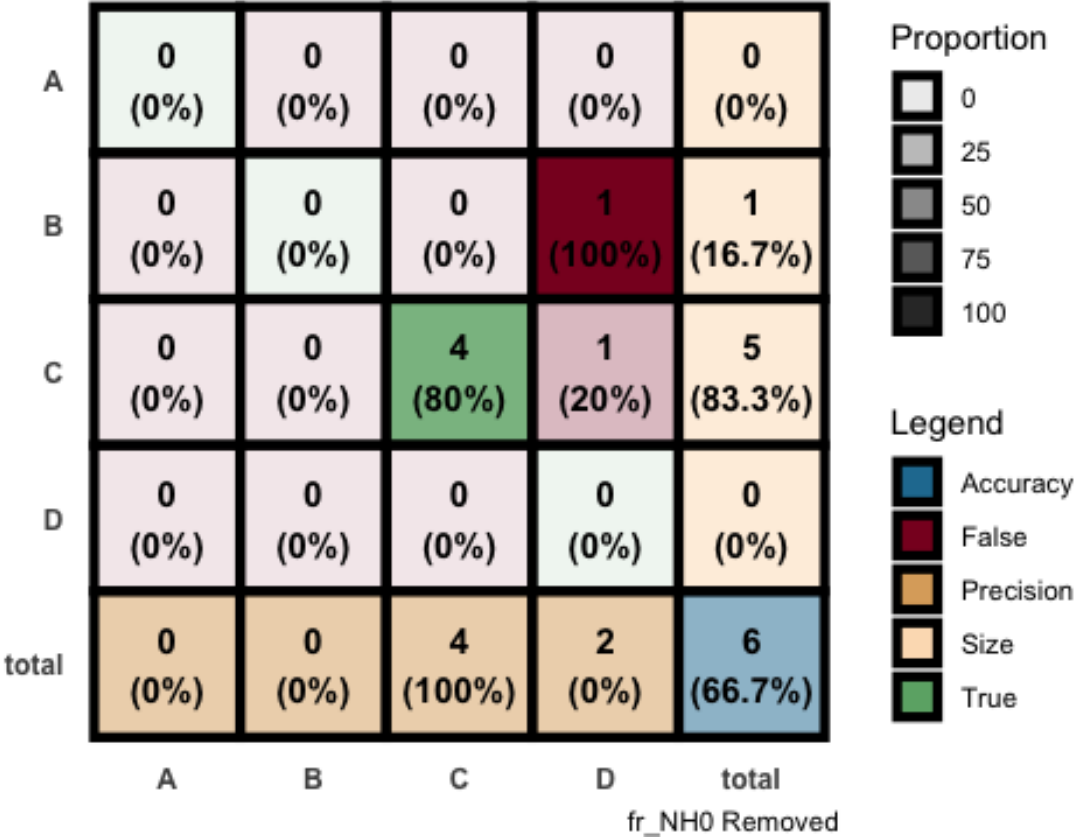
## Test Set

### Probability Heatmap



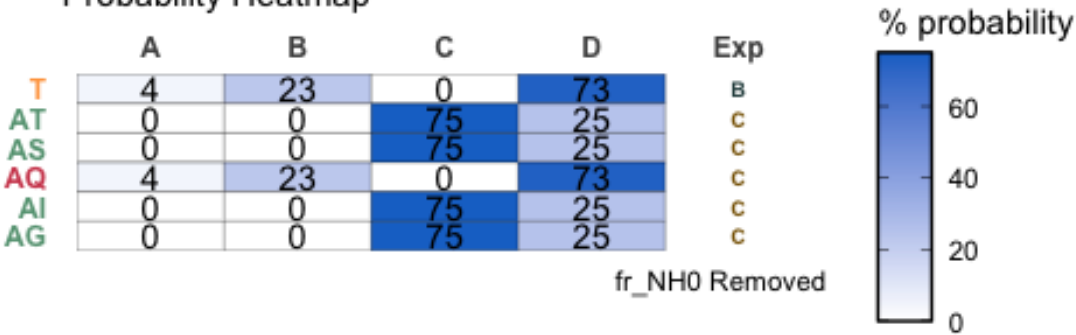
External Validation

Confusion Matrix



External Validation

Probability Heatmap



## 2. fr\_aldehyde Removed

formula

class ~ fr\_NH0 + fr\_amide + fr\_pyridine

Full Model Stats - Overall Accuracy and Pseudo-R2

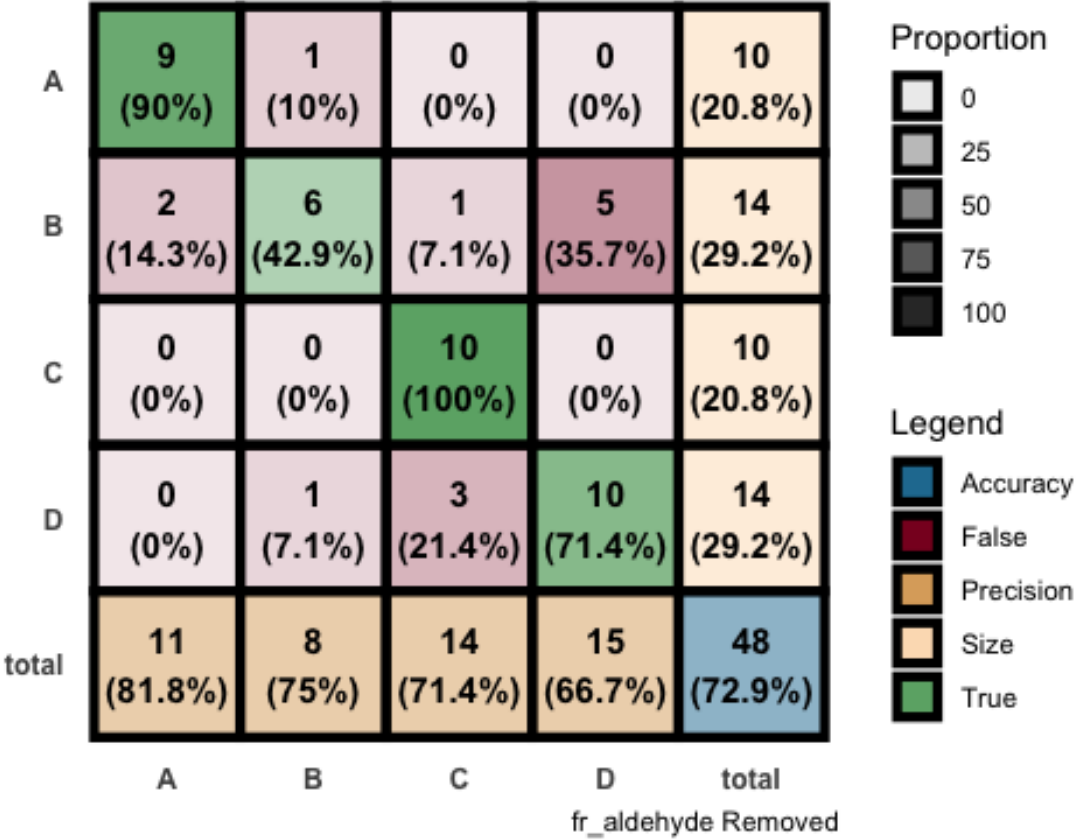
Accuracy    McFadden\_R2

72.92%    0.54

Model Coefficients

	(Intercep t)	fr_NH0 1	fr_NH0 2	fr_NH05	fr_amid e1	fr_amid e2	fr_pyridin e1	fr_pyridin e2
2	34.5623 88	- 5.1248 64	- 33.463 78	- 13.3015 97	- 14.2707 7	- 18.9664 0	-30.82382	22.84660 9
3	- 4.24679 5	- 6.3453 93	- 15.303 82	- 1.36550 2	27.1614 8	47.3129 1	-45.92722	-1.138872
4	35.0001 20	- 6.6828 34	- 35.000 13	18.0909 48	- 13.1182 8	- 21.0878 4	-52.78198	- 10.53799 6

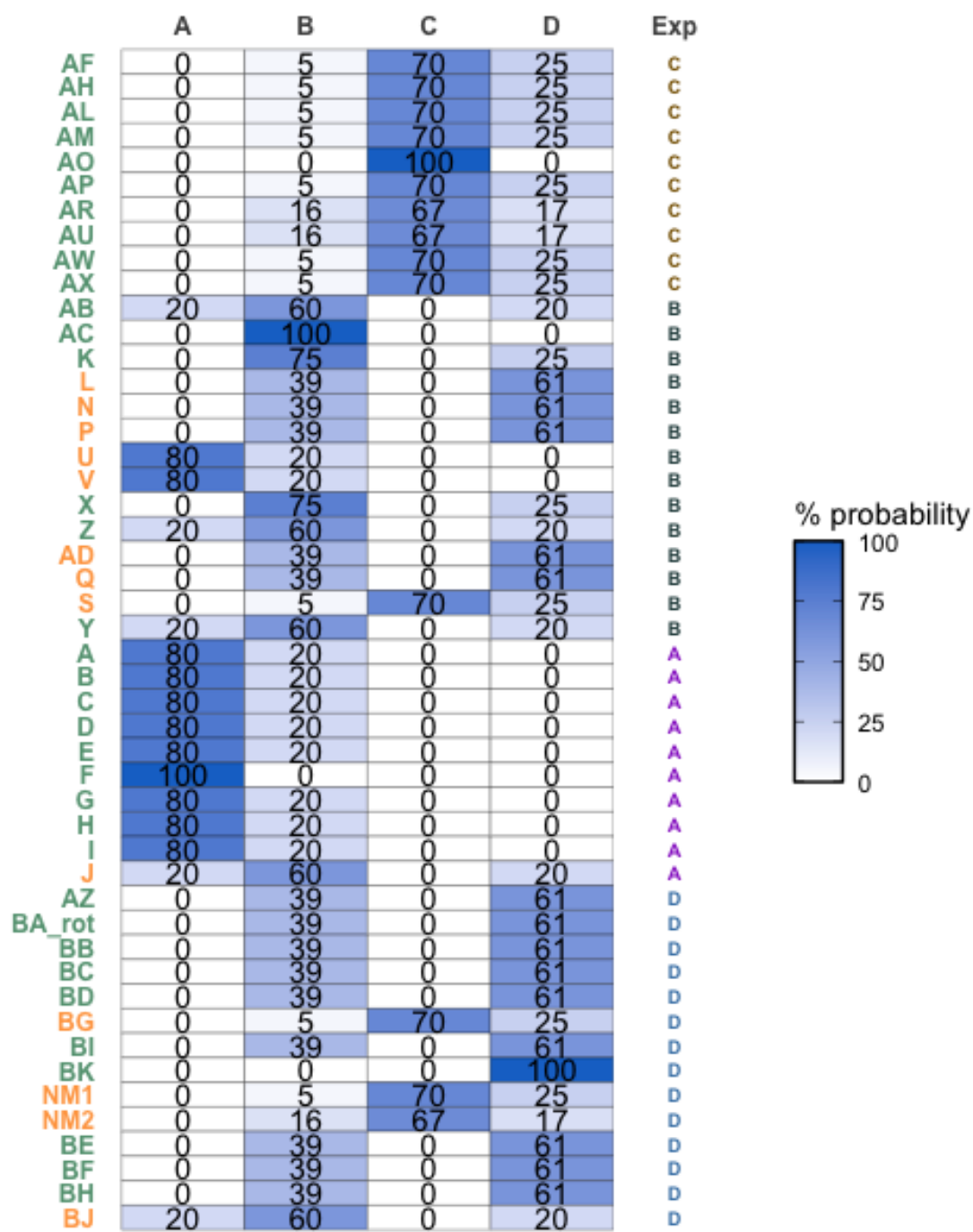
Training Set  
Confusion Matrix





# Training Set

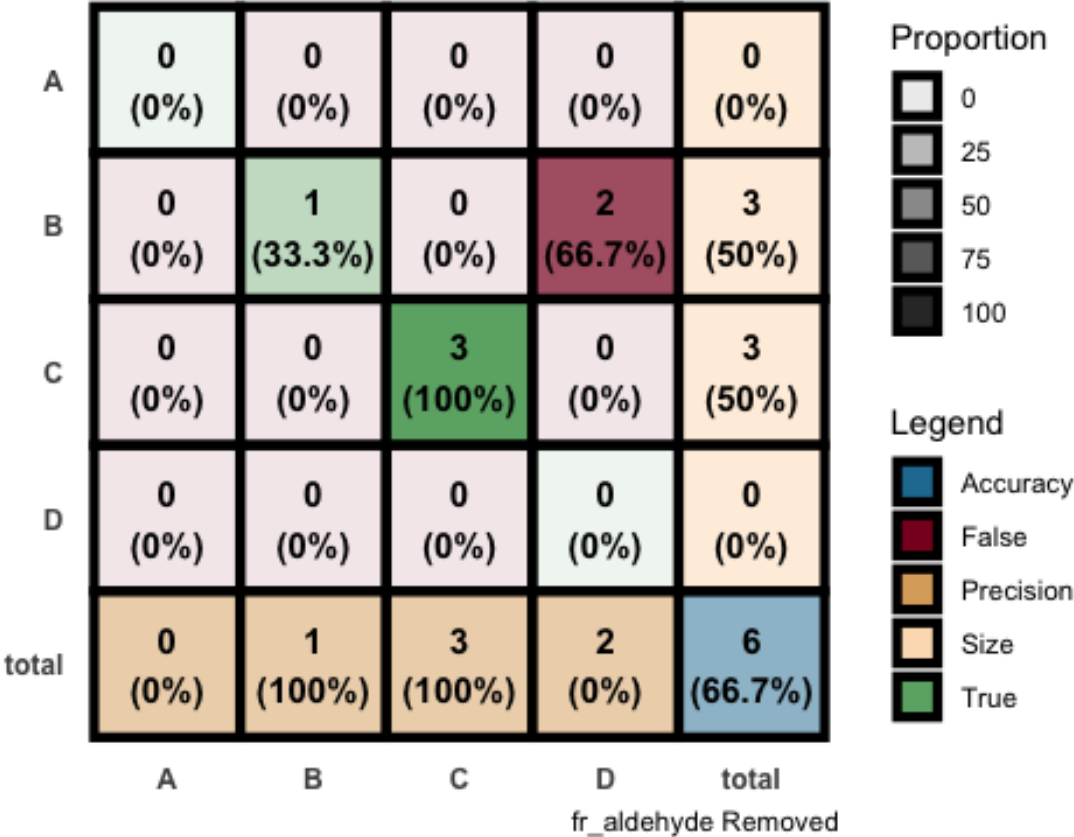
## Probability Heatmap



fr\_aldehyde Removed

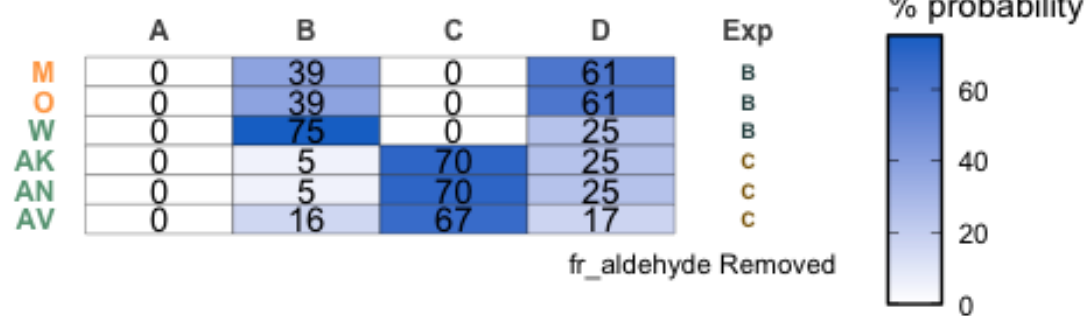
Test Set

Confusion Matrix



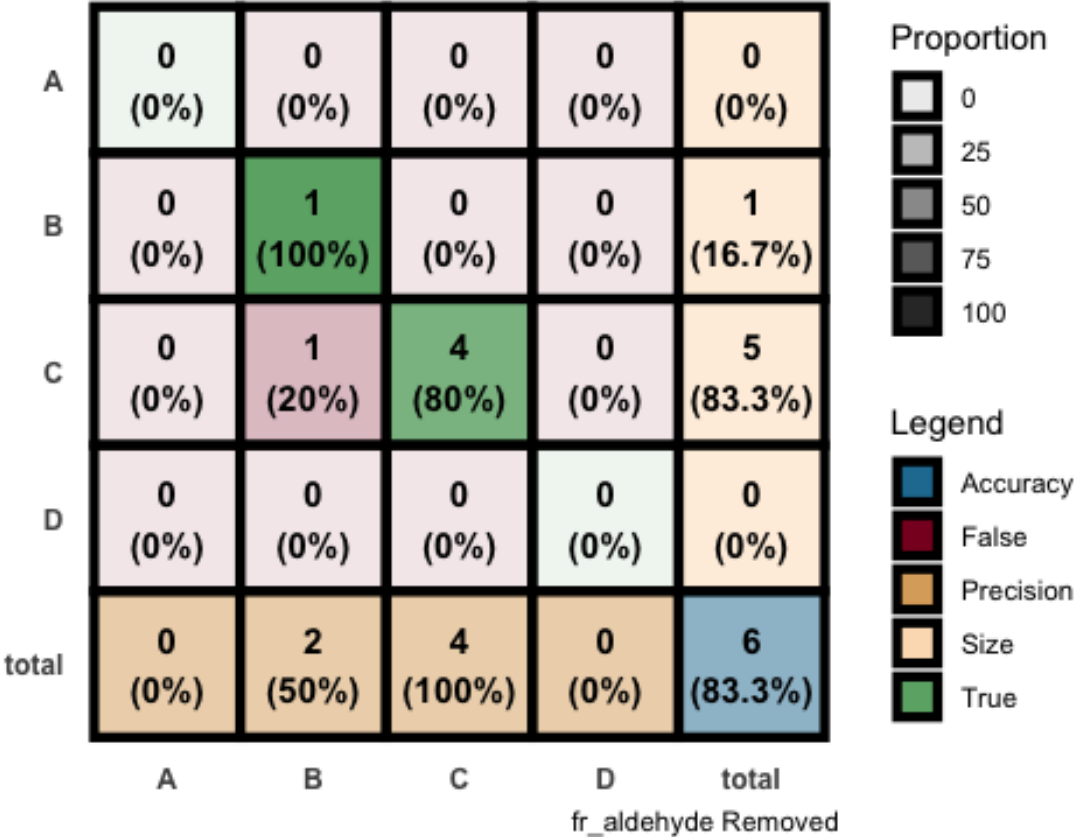
## Test Set

### Probability Heatmap



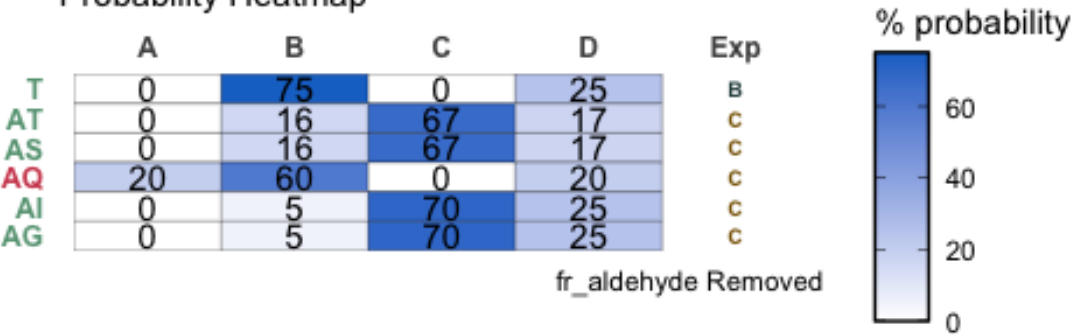
External Validation

Confusion Matrix



External Validation

Probability Heatmap



### 3. fr\_amide Removed

formula

class ~ fr\_NH0 + fr\_aldehyde + fr\_pyridine

Full Model Stats - Overall Accuracy and Pseudo-R2

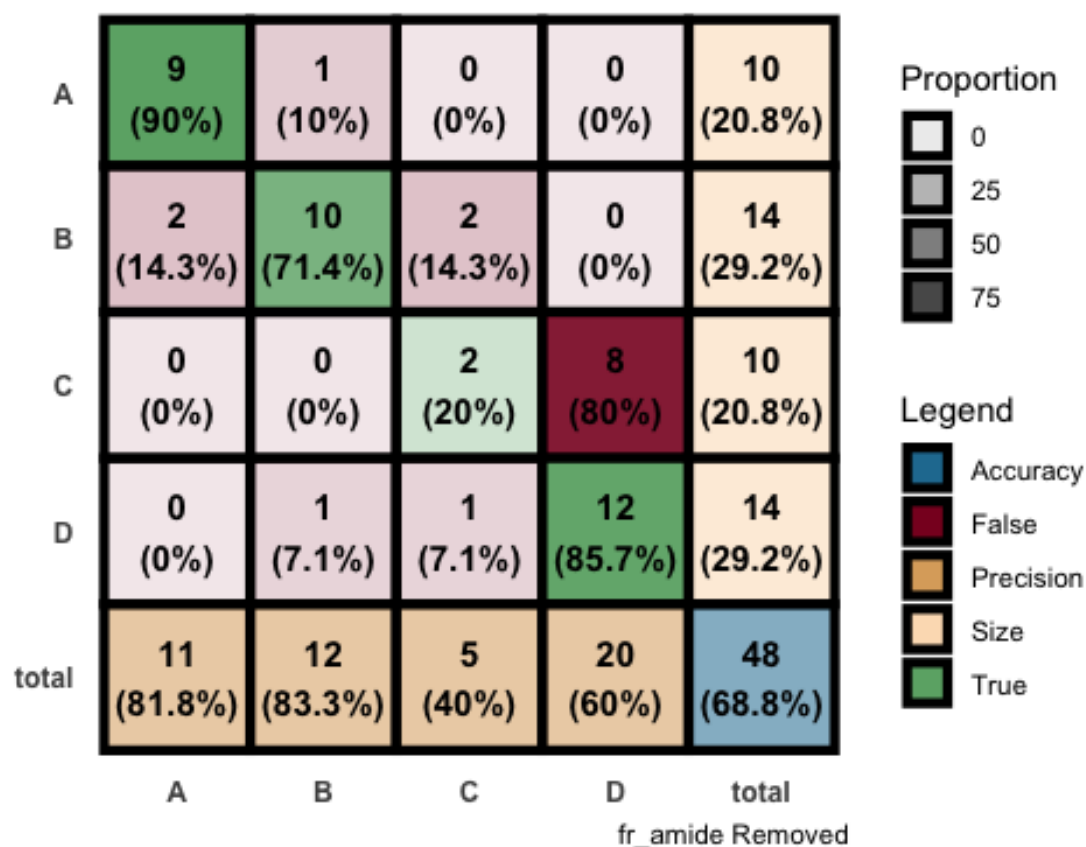
Accuracy    McFadden\_R2

68.75%    0.584

Model Coefficients

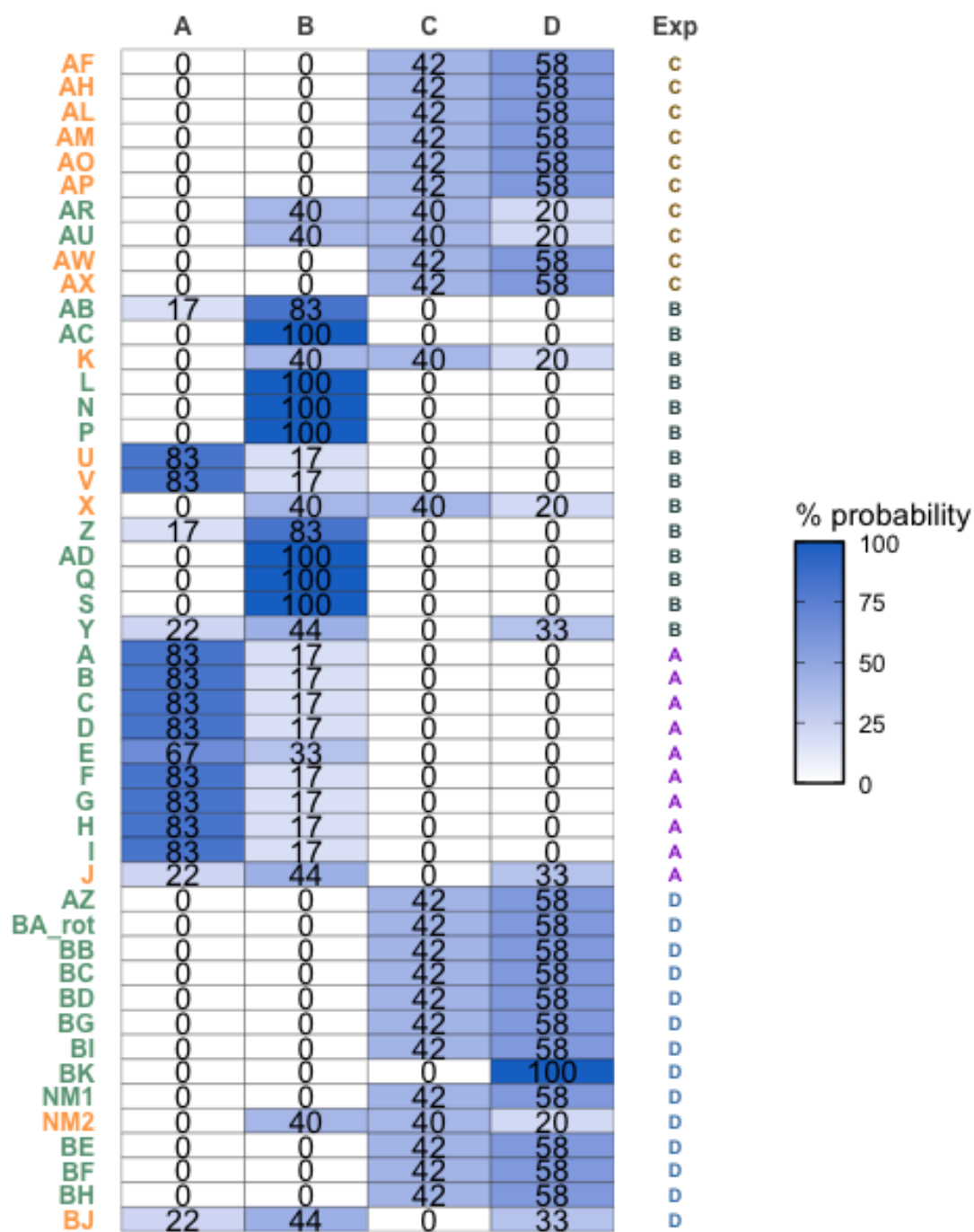
	(Intercept )	fr_NH01	fr_NH02	fr_NH05	fr_aldehyde 1	fr_pyridine 1	fr_pyridine 2
2	22.90927	16.19400 8	- 22.2162 1	-3.56486	0.9163724	-40.71274	29.776318
3	40.56576	-1.462439	- 63.2280 7	- 15.8641 5	- 50.3003813	-61.60598	-6.107742
4	40.88420	-2.474048	- 40.4788 3	23.8783 4	- 51.1005312	-58.05919	-15.830804

# Training Set Confusion Matrix



# Training Set

## Probability Heatmap



fr\_amide Removed



## Test Set

### Confusion Matrix

A	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
B	0 (0%)	2 (66.7%)	1 (33.3%)	0 (0%)	3 (50%)
C	0 (0%)	0 (0%)	1 (33.3%)	2 (66.7%)	3 (50%)
D	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
total	0 (0%)	2 (100%)	2 (50%)	2 (0%)	6 (50%)
	A	B	C	D	total
	fr_amide Removed				

### Legend

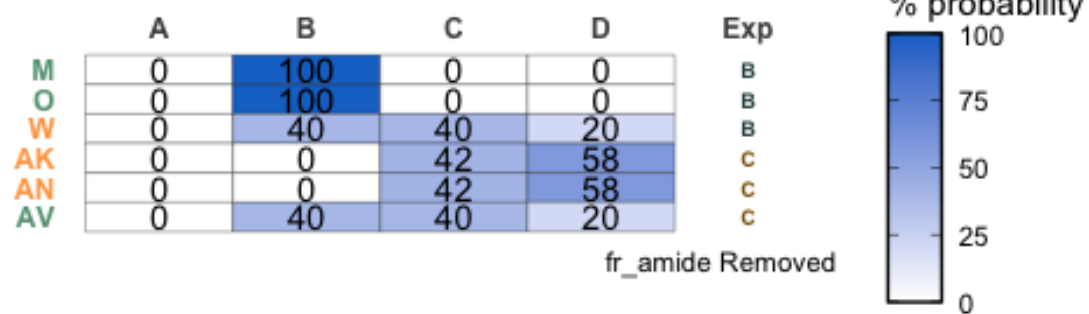
- Accuracy
- False
- Precision
- Size
- True

### Proportion

- 0
- 20
- 40
- 60

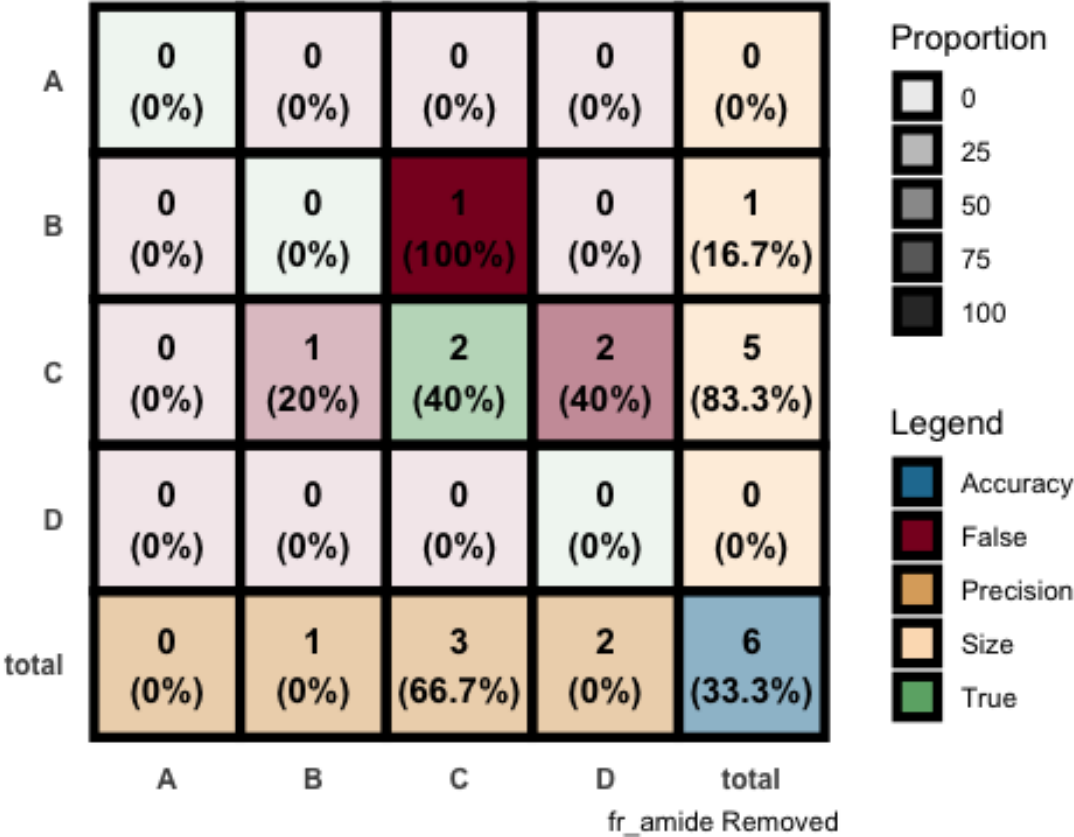
## Test Set

### Probability Heatmap



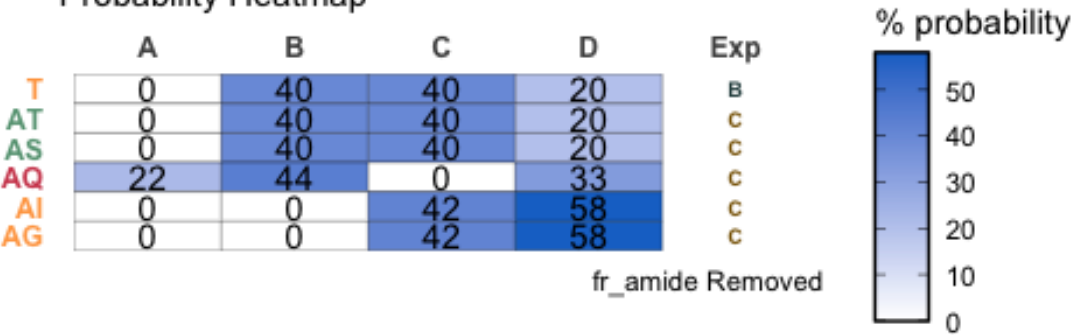
External Validation

Confusion Matrix



External Validation

Probability Heatmap



#### 4. fr\_pyridine Removed

formula

class ~ fr\_NH0 + fr\_aldehyde + fr\_amide

Full Model Stats - Overall Accuracy and Pseudo-R2

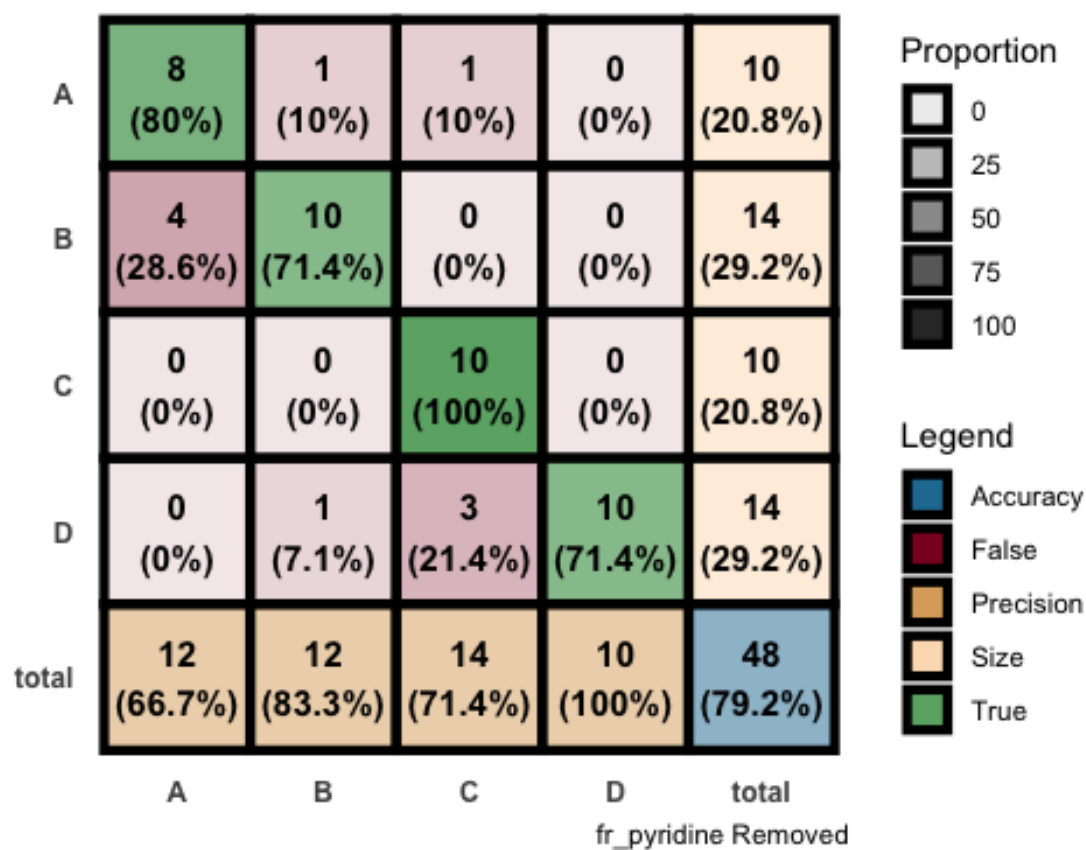
Accuracy    McFadden\_R2

79.17%    0.676

Model Coefficients

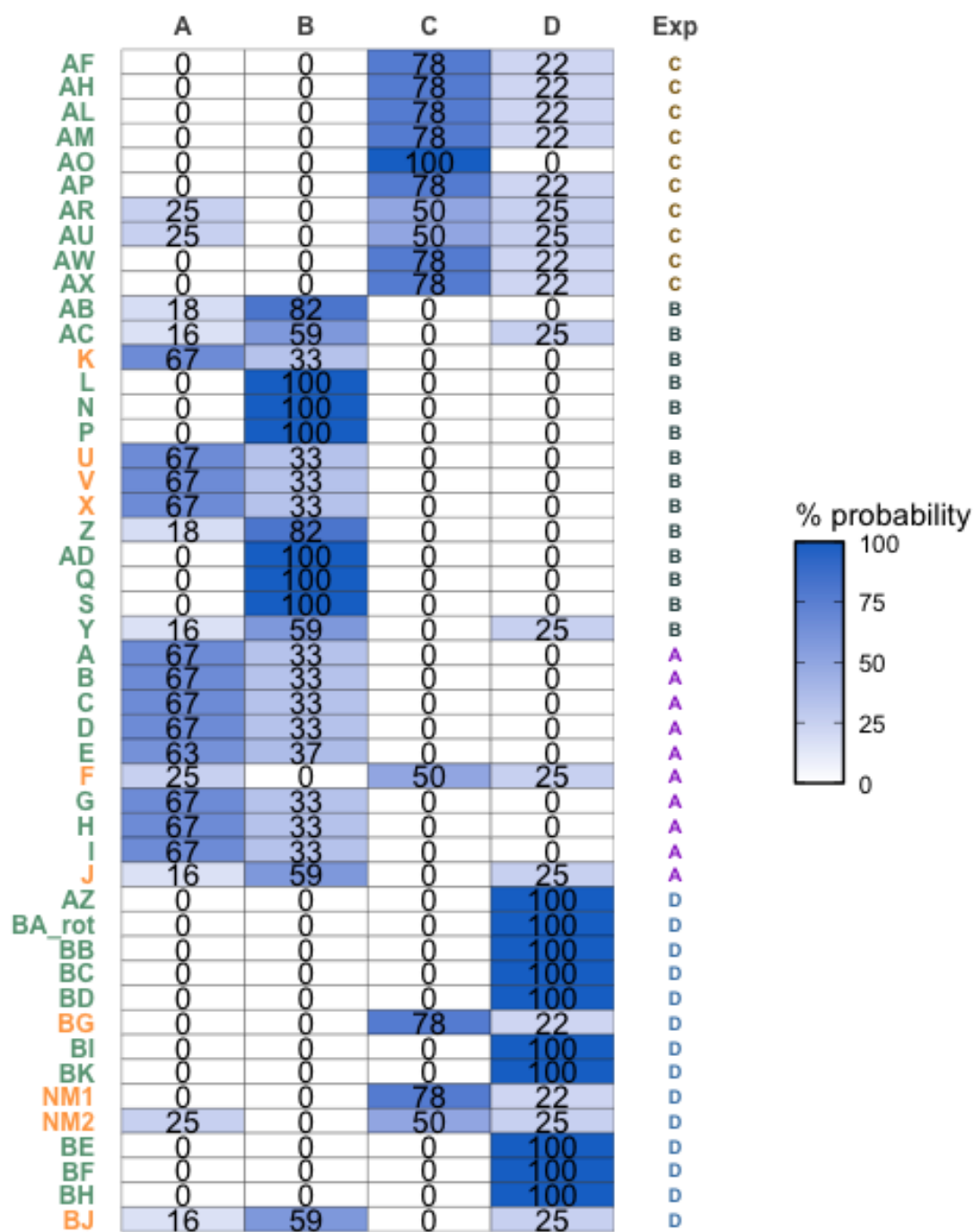
	(Intercept )	fr_NH01	fr_NH02	fr_NH05	fr_aldehyde1	fr_amide 1	fr_amide2
2	40.004467	- 40.7119 1	- 38.6832 9	-6.192403	0.1676057	- 22.71663	-9.730619
3	9.550755	- 85.3334 7	- 25.6255 9	-2.528876	-95.4630660	76.47584	60.01111 9
4	58.940766	- 84.7738 7	- 58.4817 7	16.21055 2	- 104.975011 1	25.83309	- 36.68719 0

# Training Set Confusion Matrix



# Training Set

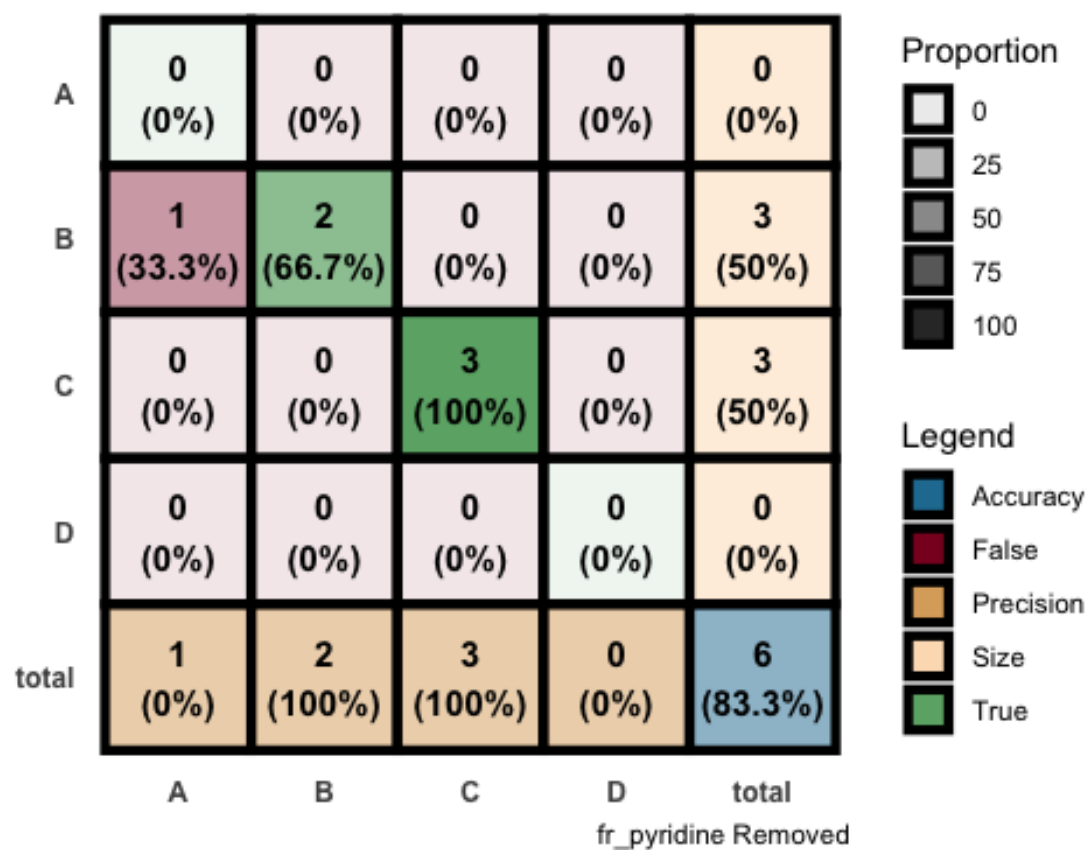
## Probability Heatmap



fr\_pyridine Removed

## Test Set

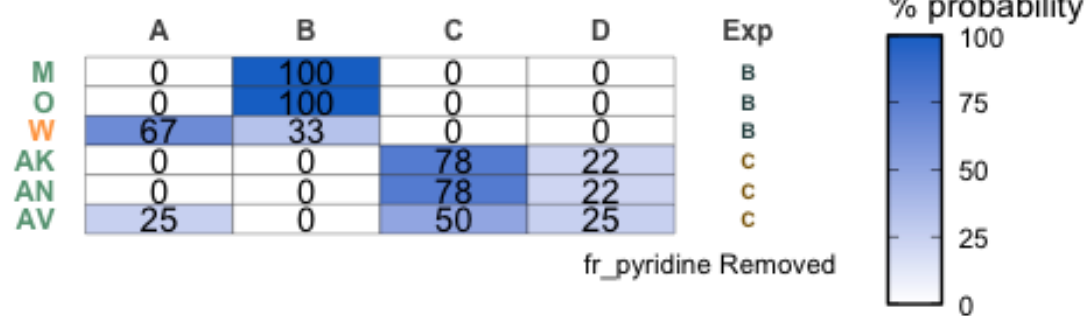
### Confusion Matrix





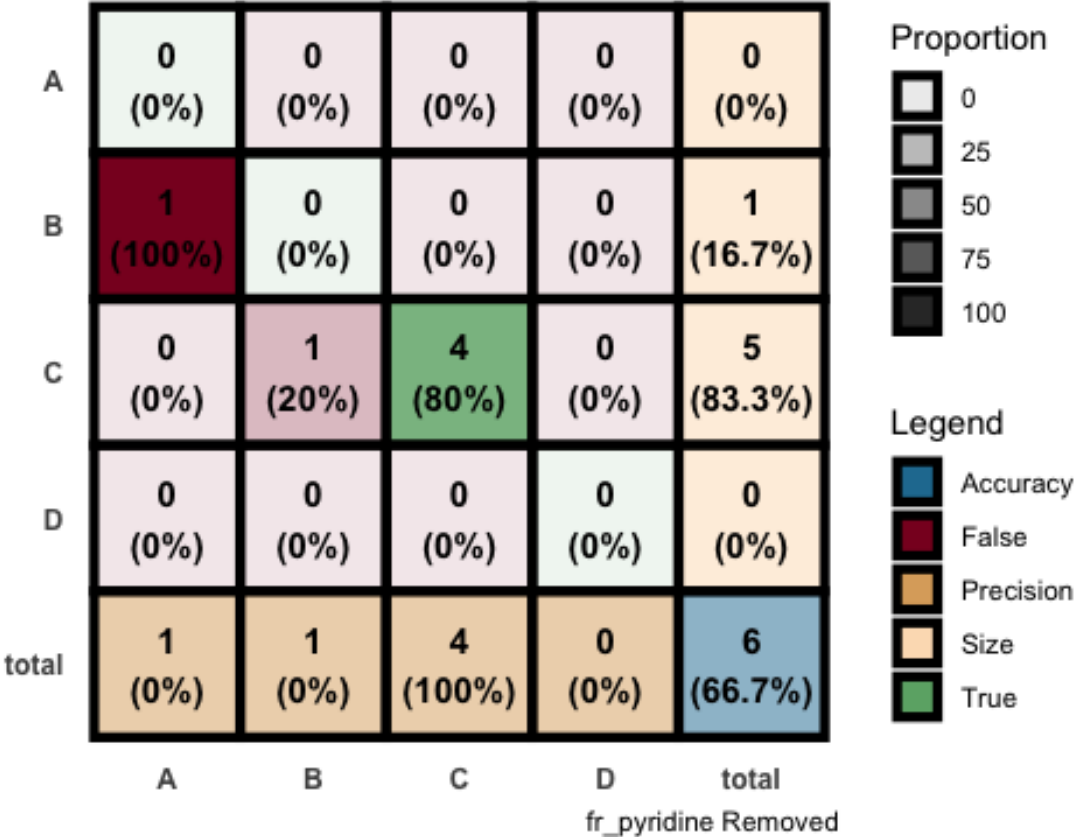
## Test Set

### Probability Heatmap



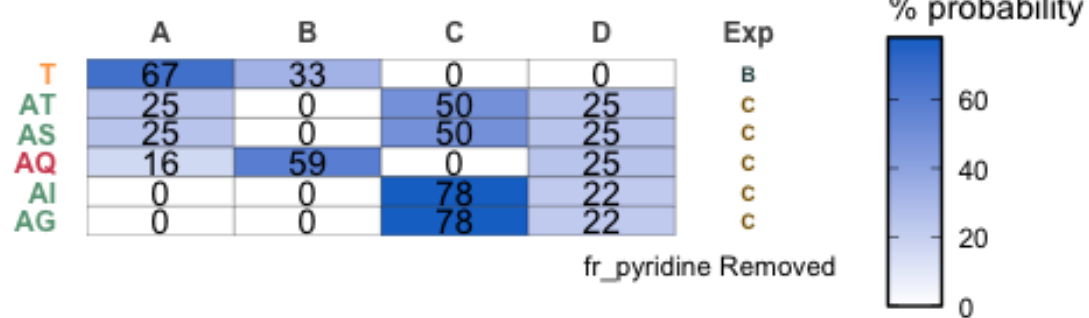
External Validation

Confusion Matrix



## External Validation

### Probability Heatmap



## Section Results Summary

	Training Accuracy	Test Accuracy
Full Model	85.42	100.0
1. fr_NH0 Removed	81.25	83.3
2. fr_aldehyde Removed	72.92	66.7
3. fr_amide Removed	68.75	50.0
4. fr_pyridine Removed	79.17	83.3