

Aldehyde Deuteration - Best Model Feature Analysis

Table of Contents

Model Search.....	1
1st Place.....	1
Model Analysis - Variable Impact.....	8
1. NPA_1 Removed	8
2. Dist(2, 7) Removed	15
3. dip_y Removed.....	22
4. -2-3- Removed.....	29

Model Search

1st Place

Table:

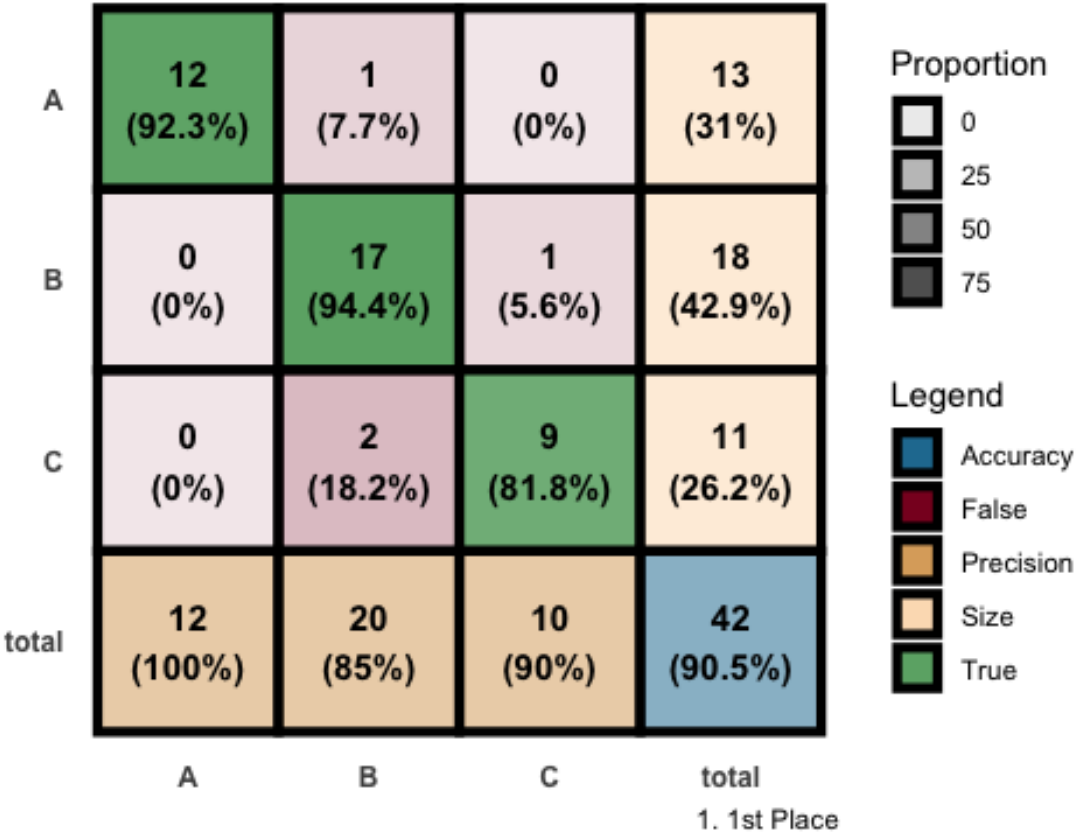
Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
90.48%	0.577

Model Coefficients

	x
-2-3-	-0.3801322
dip_y	-1.5680528
Dist(2, 7)	295.8494658
NPA_1	47.5647077

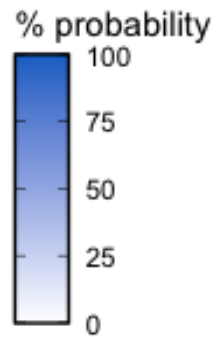
Training Set
Confusion Matrix



Training Set

Probability Heatmap

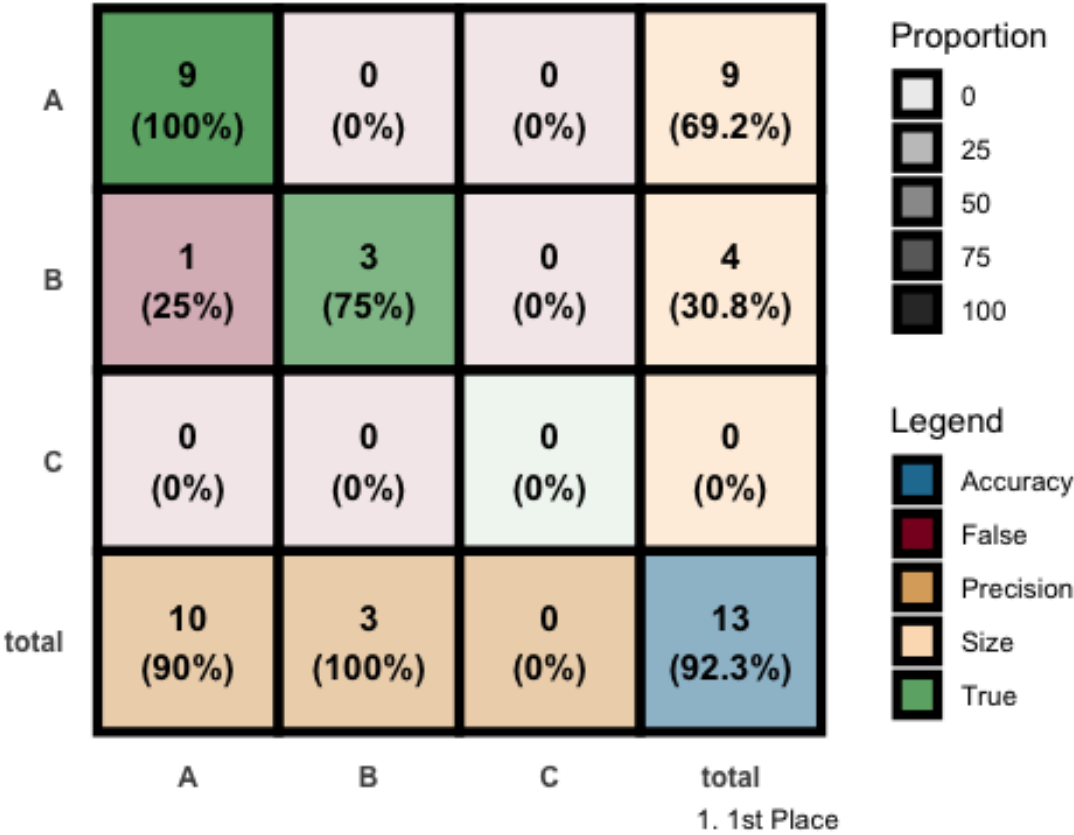
	A	B	C	Exp
2A	57	42	1	A
2G	98	2	0	A
2H	98	2	0	A
2L	88	11	0	A
2M	89	11	0	A
2O	100	0	0	A
2R	13	80	6	A
2S	71	28	0	A
2T	74	25	0	A
2U	99	1	0	A
2V	82	18	0	A
2J	87	13	0	A
2P	98	2	0	A
2Y	2	68	30	B
2Z	2	65	33	B
2AA	1	51	48	B
2AE	10	81	8	B
2AF	46	53	1	B
2AG	18	77	4	B
2AH	16	79	5	B
2AJ	2	62	36	B
2AK	6	81	13	B
2AM	7	81	13	B
2AO	23	74	3	B
2W	27	70	3	B
2X	16	79	5	B
2AB	3	71	26	B
2AC	5	79	15	B
2AP	20	76	4	B
2AQ	22	75	4	B
2AR	0	24	76	B
2AS	1	44	55	C
2AT	1	44	55	C
2AU	38	60	2	C
2AV	1	46	54	C
2AW	5	79	16	C
2AX	0	2	98	C
2AY	0	23	77	C
2AZ	0	2	98	C
2BA	0	0	100	C
2BB	0	0	100	C
2BC	0	0	100	C



1. 1st Place

Test Set

Confusion Matrix



Test Set

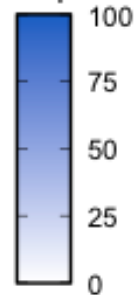
Probability Heatmap

	A	B	C
2B	86	14	0
2C	85	15	0
2D	100	0	0
2E	83	17	0
2F	81	19	0
2I	92	8	0
2K	100	0	0
2N	77	23	0
2Q	98	2	0
2AD	29	69	2
2AI	5	78	17
2AL	67	33	1
2AN	10	81	8

Exp

A
A
A
A
A
A
A
A
A
B
B
B
B

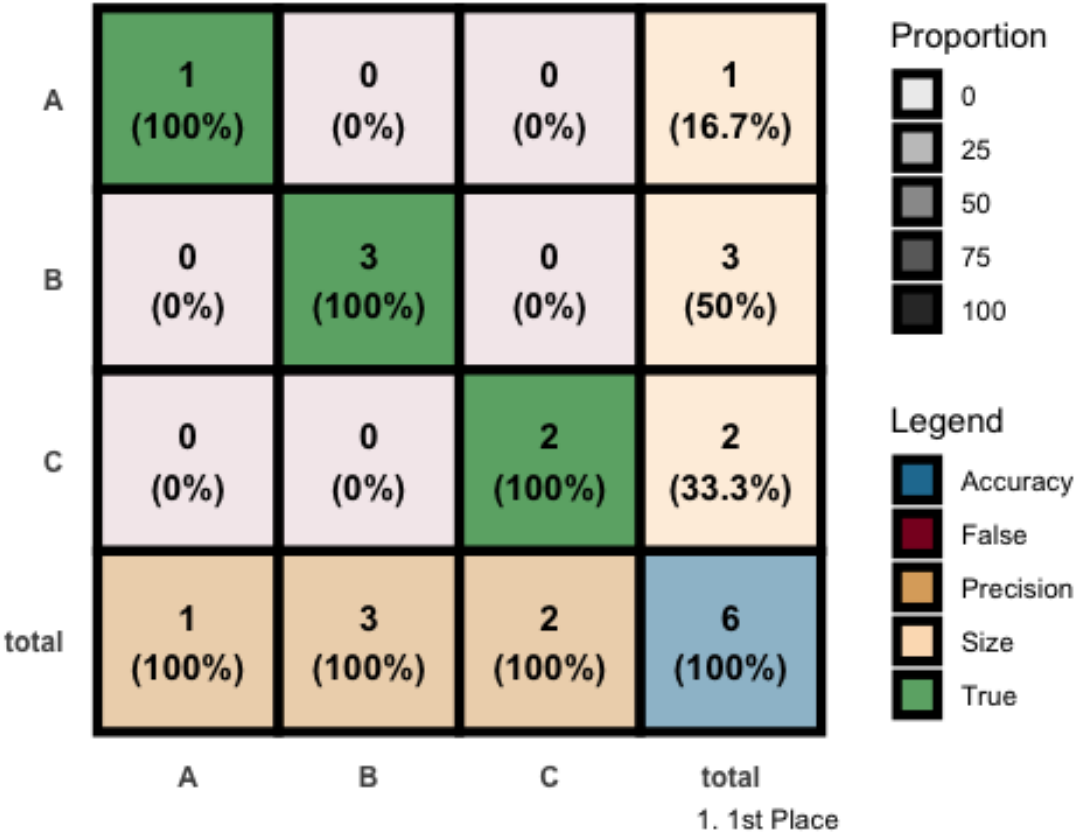
% probability



1. 1st Place

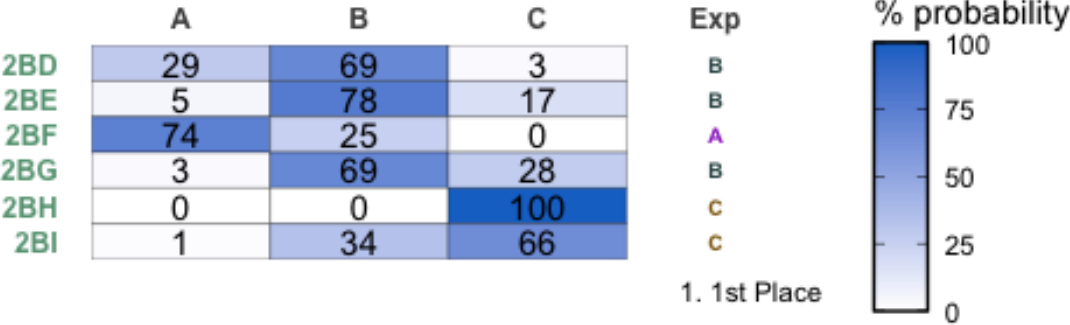
External Validation

Confusion Matrix



External Validation

Probability Heatmap



New Molecules Prediction

New Molecules - Prediction of Conditions

	Predicted	1	2	3
2BJ	3	0	0	100
2BK	1	73	27	0
2BL	2	29	69	3
2BM	1	100	0	0
2BN	3	0	0	100
2BO	3	0	0	100
2BP	1	100	0	0
2BQ	1	100	0	0
2BR	1	100	0	0
2BS	1	100	0	0
2BT	1	100	0	0
2BU	1	57	42	1
2BV	2	25	72	3
2BW	2	1	58	41

Model Analysis - Variable Impact

1. NPA_1 Removed

formula

```
class ~ -2-3- + dip_y + Dist(2, 7)
```

Full Model Stats - Overall Accuracy and Pseudo-R2

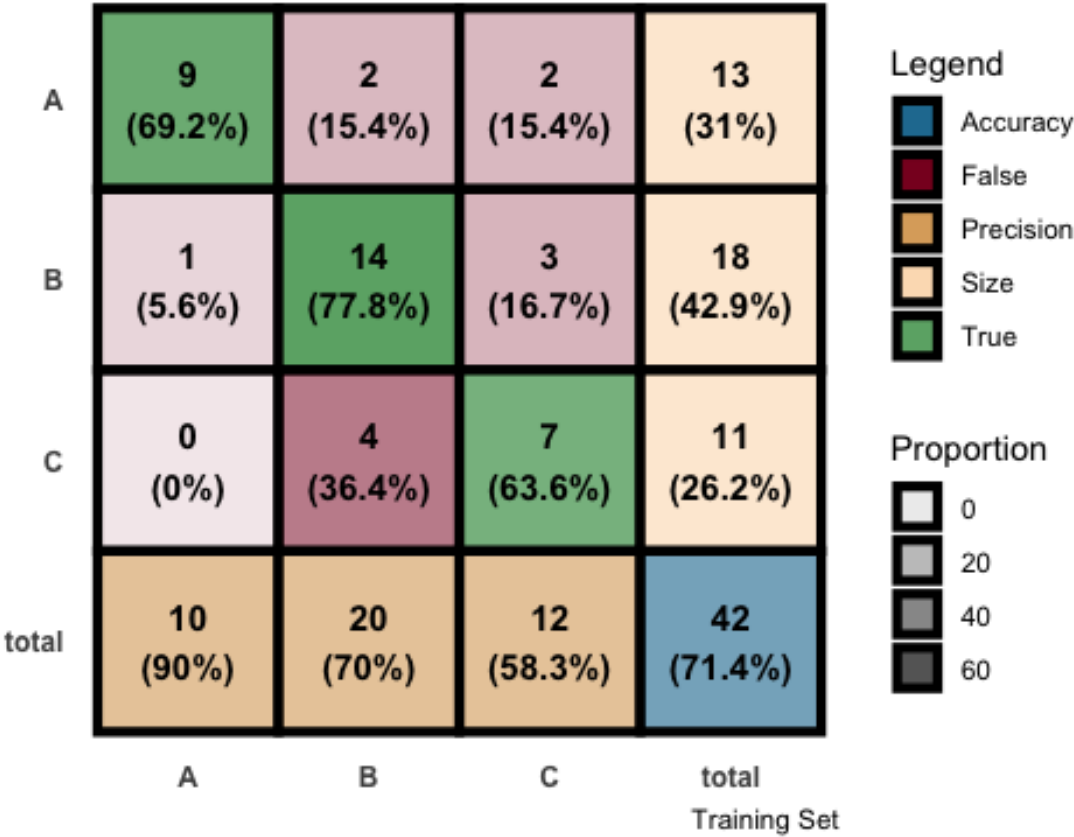
Accuracy	McFadden_R2
71.43%	0.338

Model Coefficients

	x
-2-3-	-0.1882832
dip_y	-0.9381293
Dist(2, 7)	145.2808583

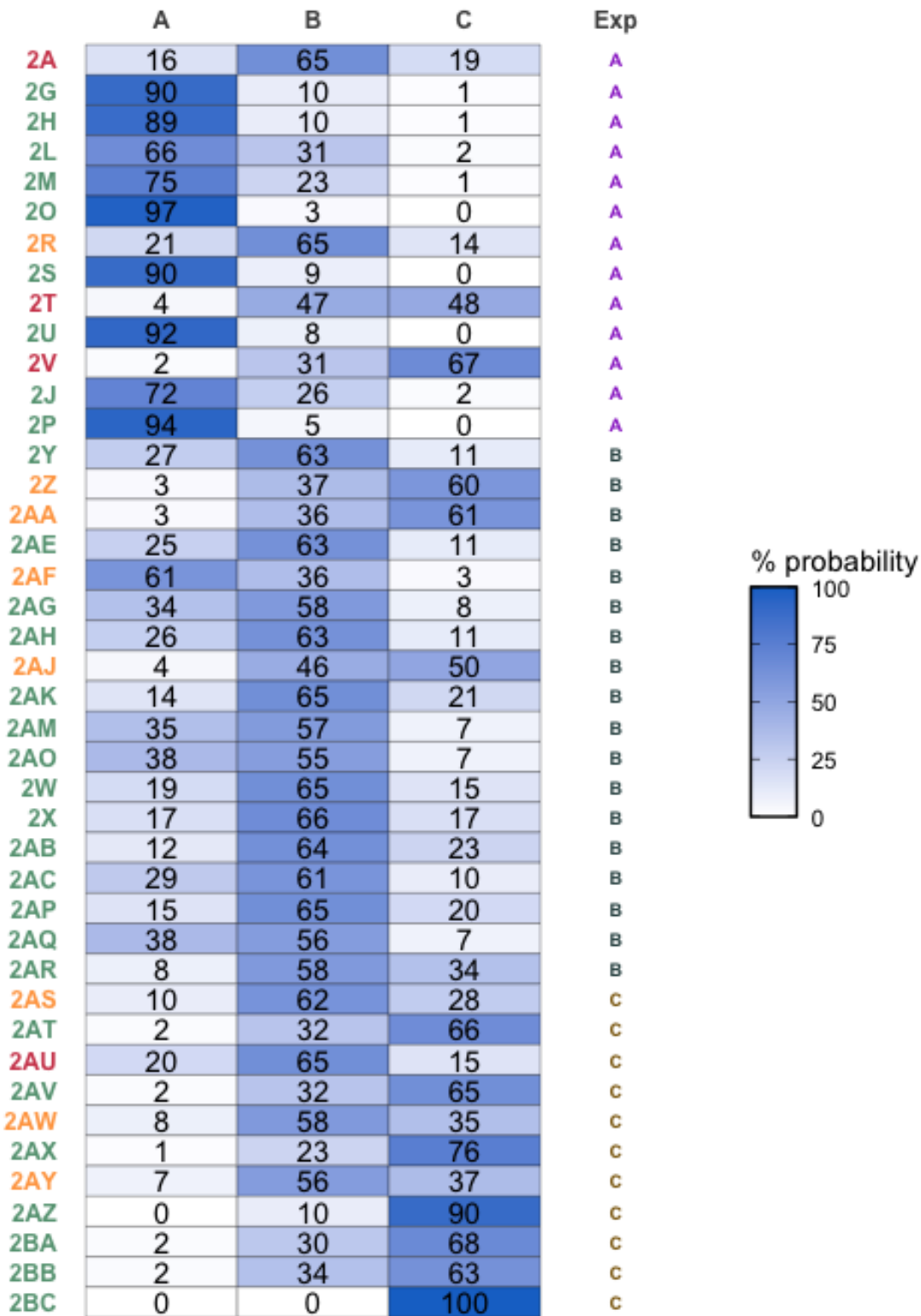
NPA_1 Removed

Confusion Matrix



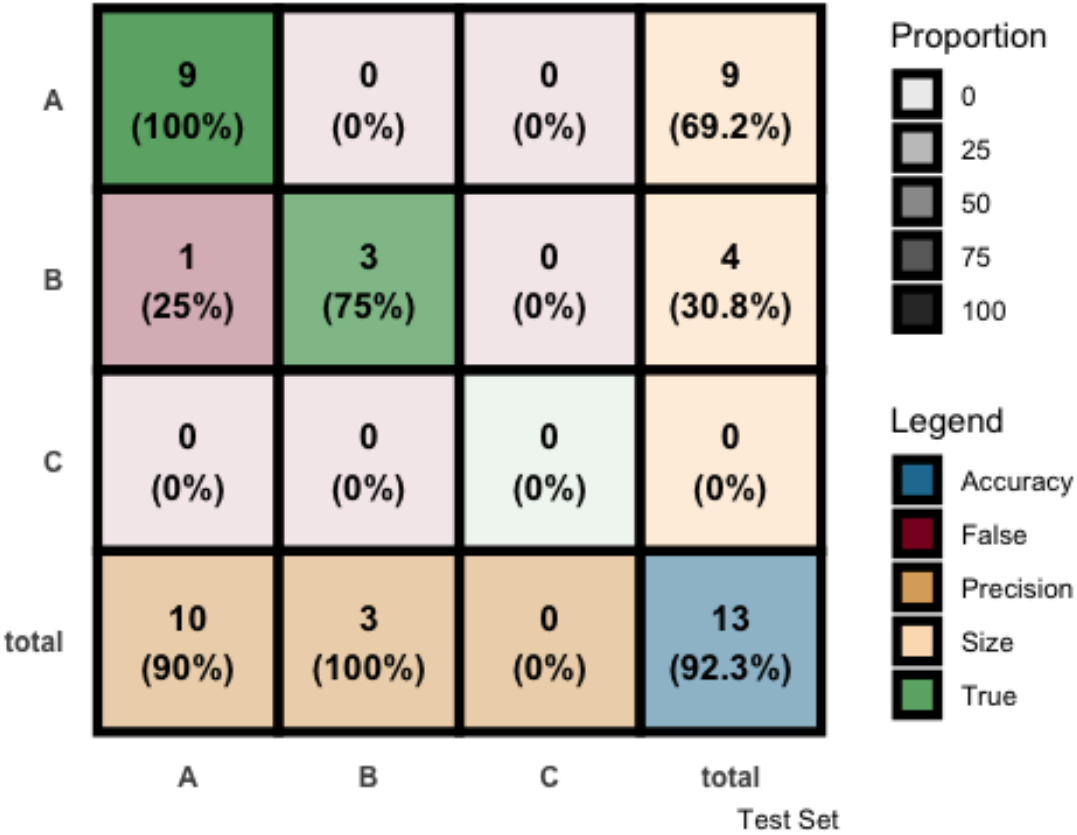
NPA_1 Removed

Probability Heatmap

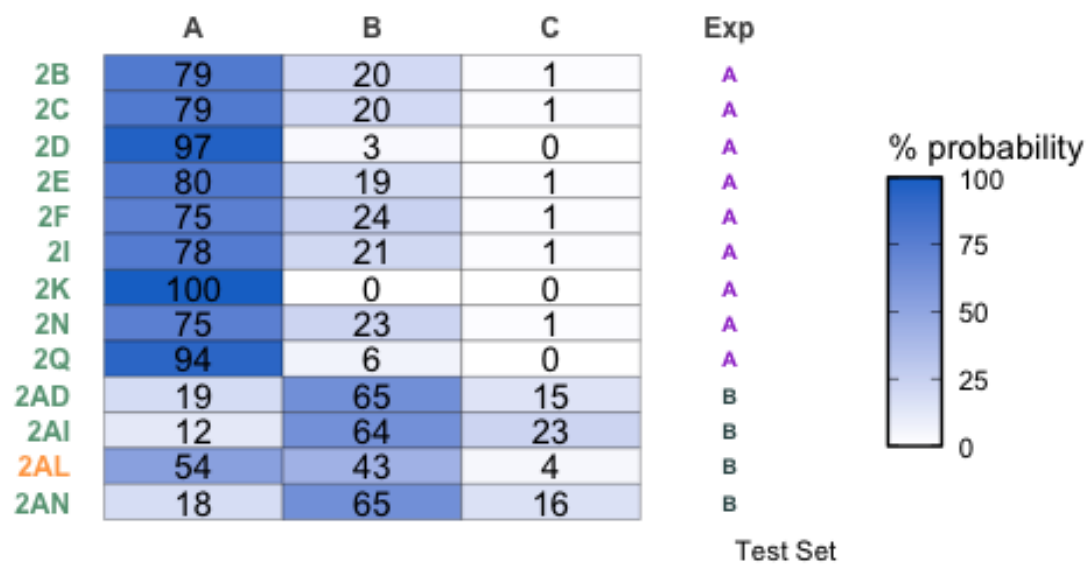


Training Set

NPA_1 Removed
Confusion Matrix

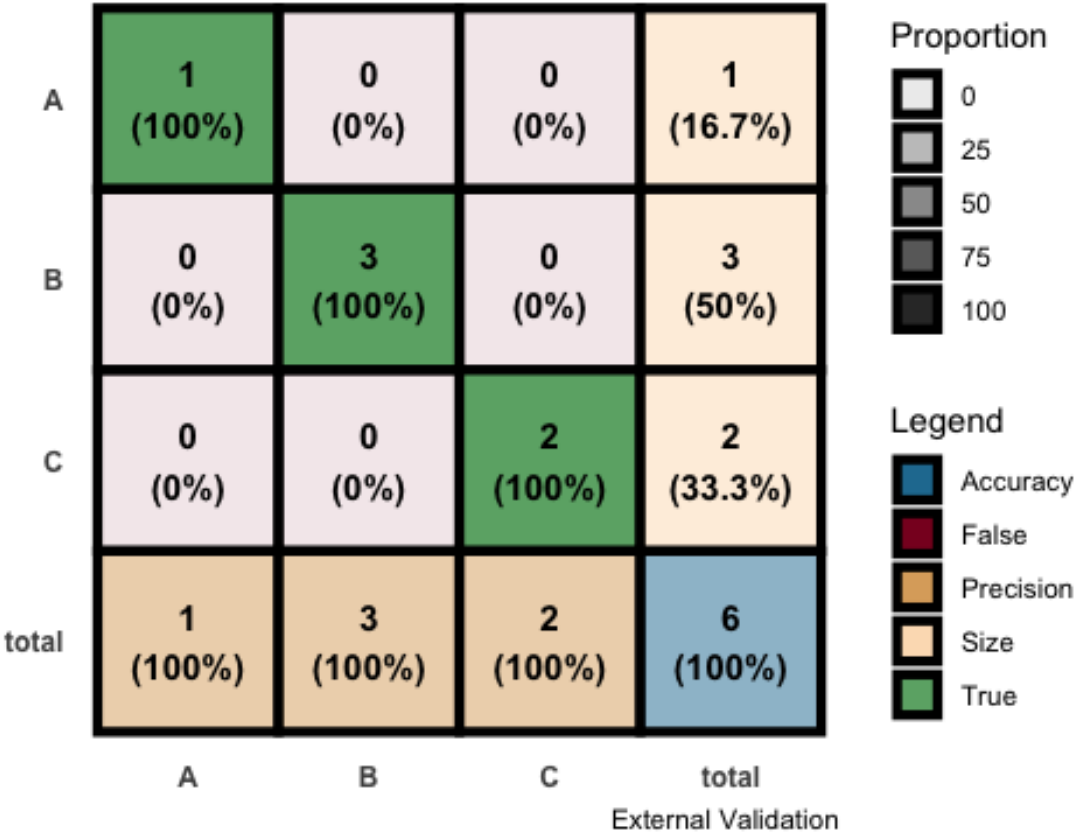


NPA_1 Removed Probability Heatmap

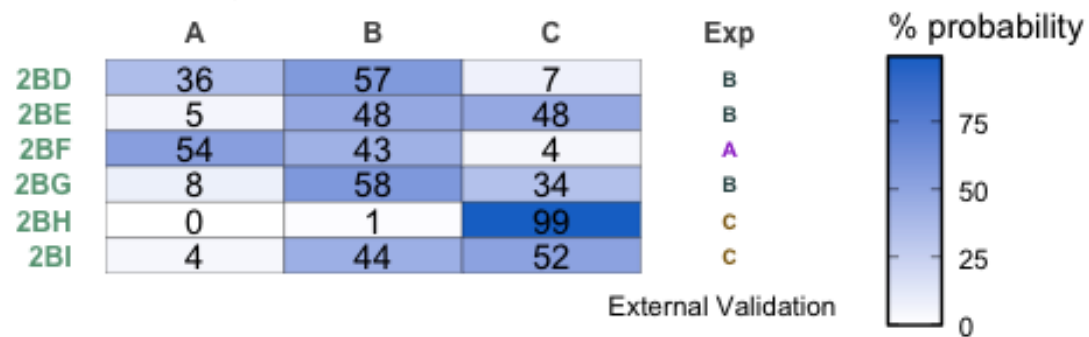


NPA_1 Removed

Confusion Matrix



NPA_1 Removed Probability Heatmap



2. Dist(2, 7) Removed

formula

class ~ -2-3- + dip_y + NPA_1

Full Model Stats - Overall Accuracy and Pseudo-R2

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

78.57%	0.452
--------	-------

Model Coefficients

	x
--	---

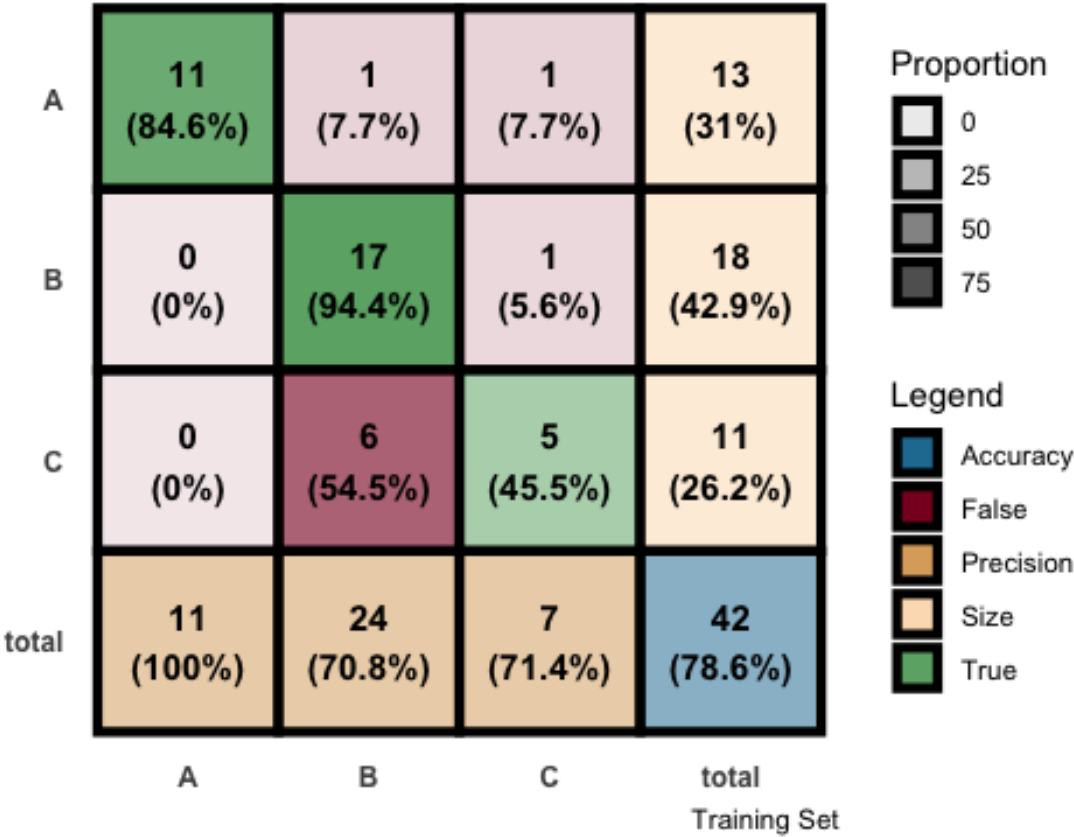
-2-3-	-0.076705
-------	-----------

dip_y	-1.830226
-------	-----------

NPA_1	27.040639
-------	-----------

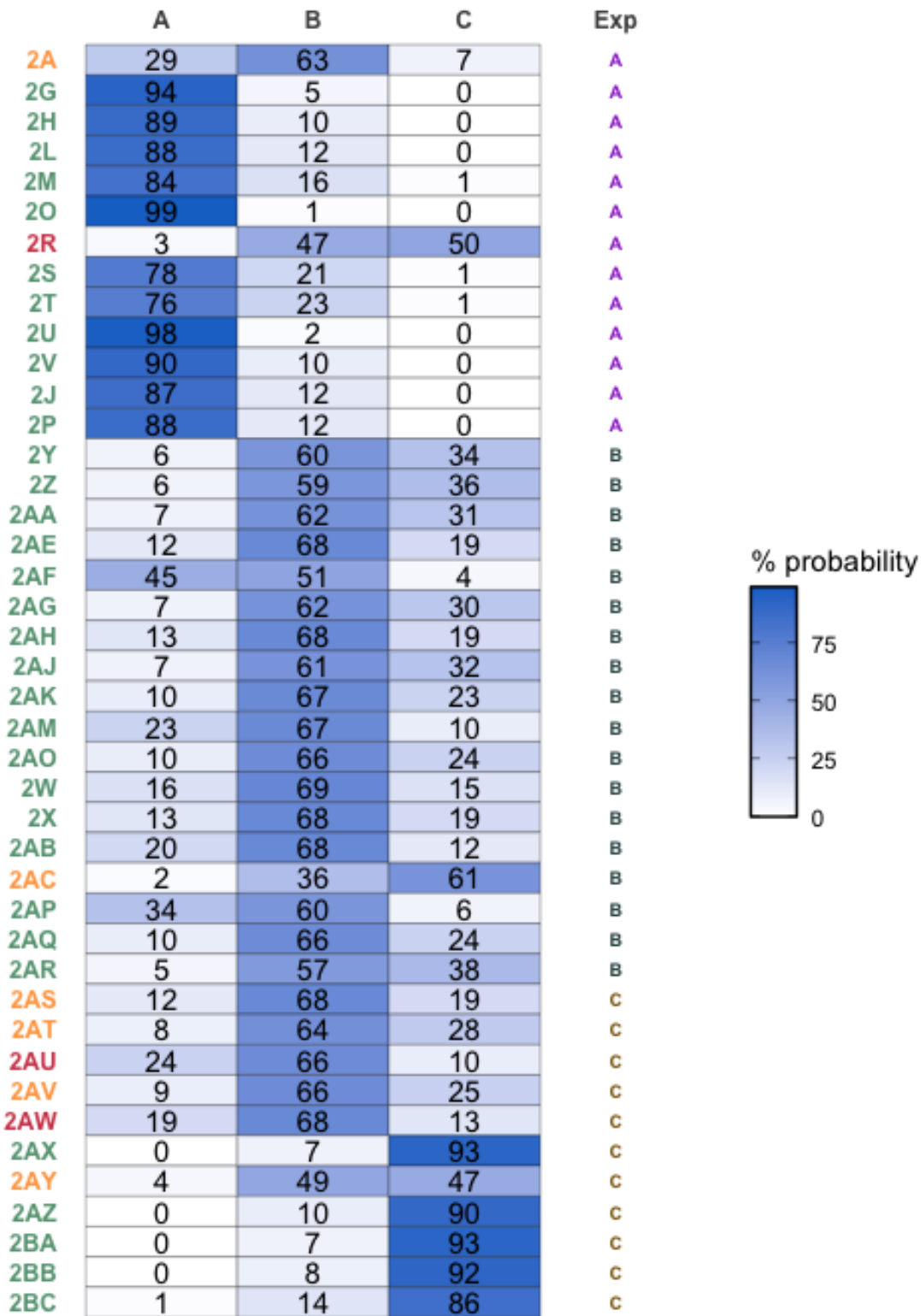
Dist(2, 7) Removed

Confusion Matrix



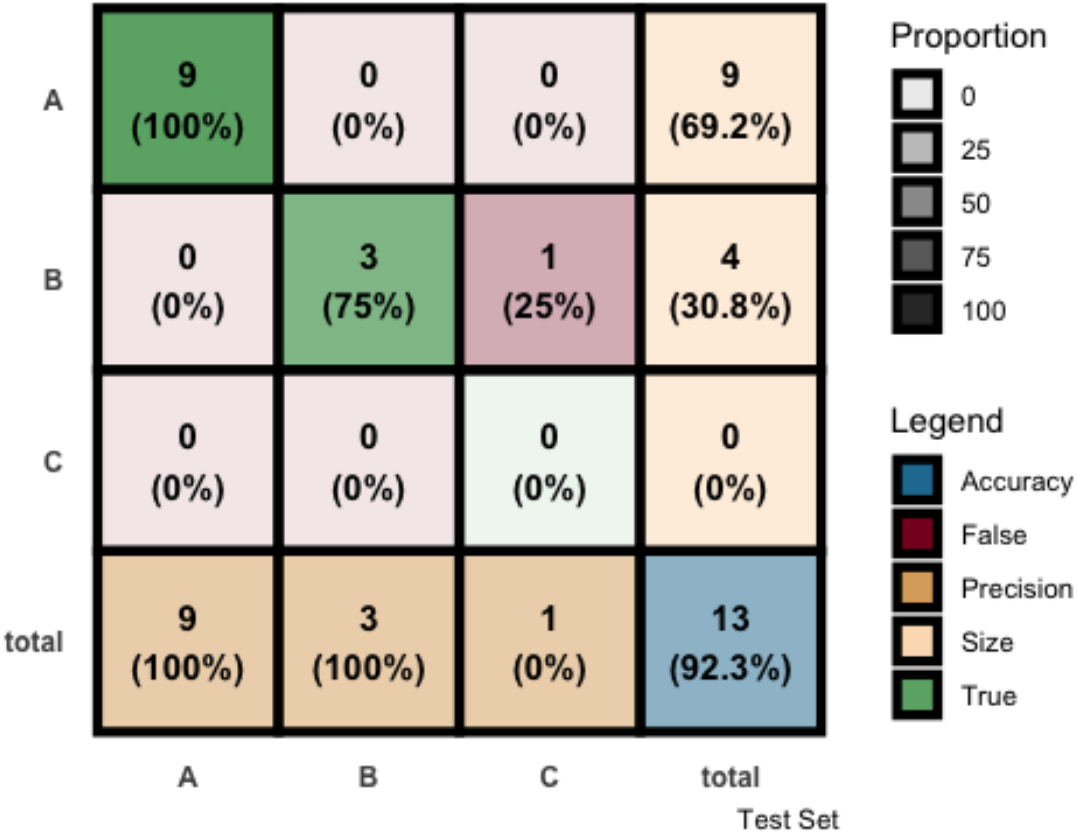
Dist(2, 7) Removed

Probability Heatmap



Training Set

Dist(2, 7) Removed
Confusion Matrix



Dist(2, 7) Removed

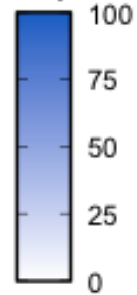
Probability Heatmap

	A	B	C
2B	64	34	2
2C	62	36	2
2D	98	2	0
2E	57	41	3
2F	58	39	2
2I	89	11	0
2K	100	0	0
2N	51	46	3
2Q	87	12	1
2AD	18	69	14
2AI	8	65	27
2AL	38	57	5
2AN	3	43	54

Exp

A
A
A
A
A
A
A
A
A
B
B
B
B

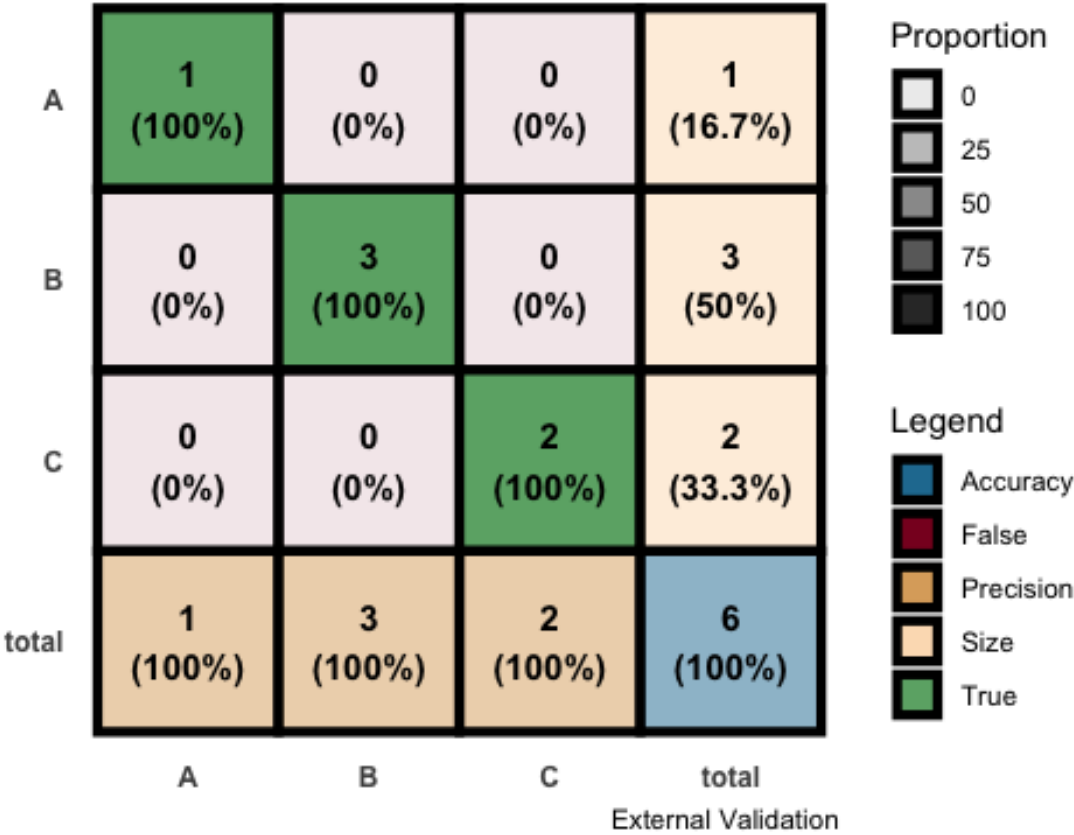
% probability



Test Set

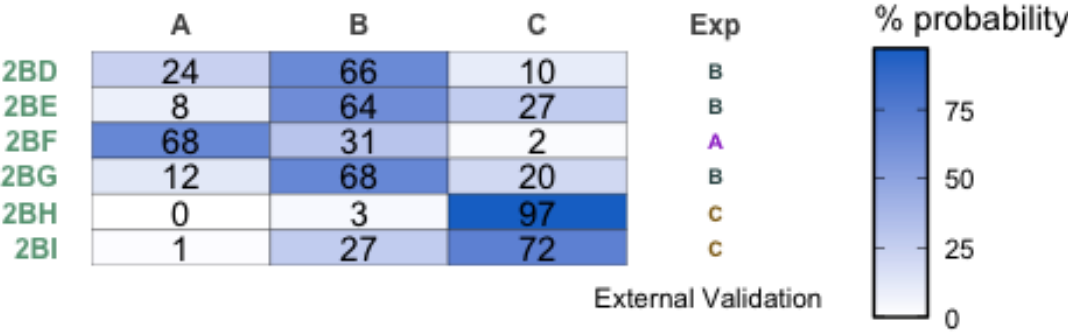
Dist(2, 7) Removed

Confusion Matrix



Dist(2, 7) Removed

Probability Heatmap



3. dip_y Removed

formula

class ~ -2-3- + Dist(2, 7) + NPA_1

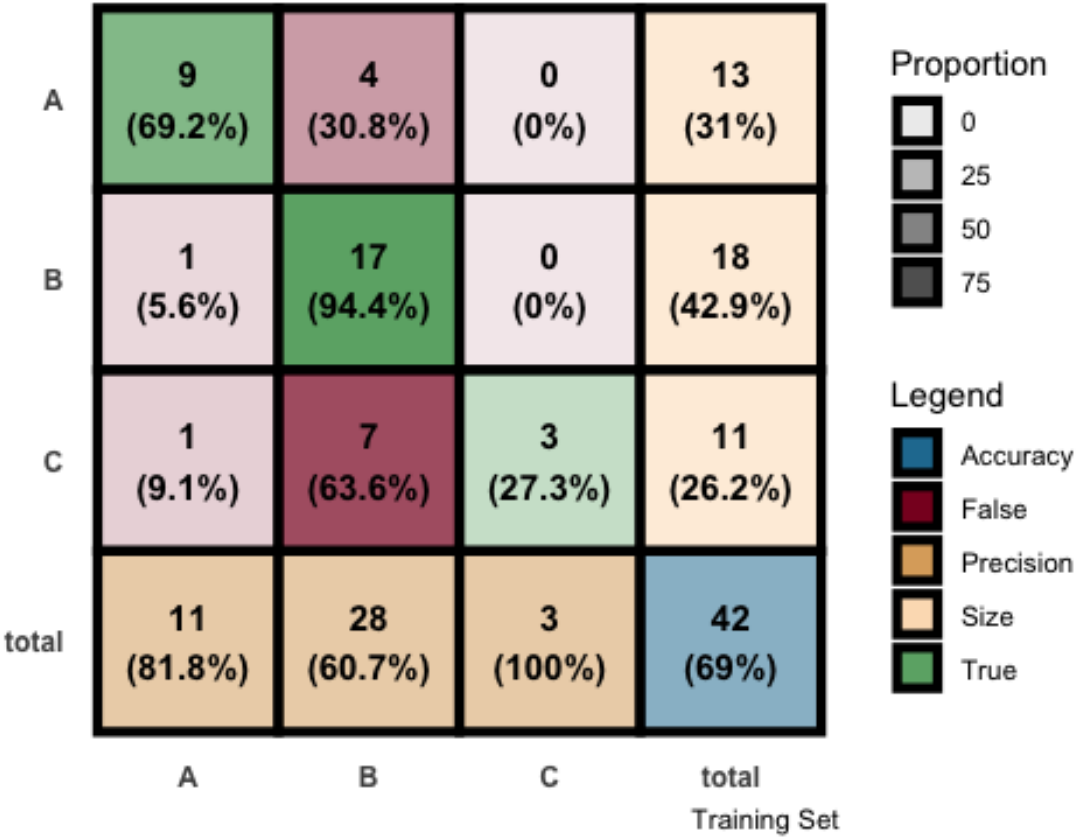
Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
69.05%	0.217

Model Coefficients

	x
-2-3-	-0.0879166
Dist(2, 7)	73.6506465
NPA_1	15.3035790

dip_y Removed
Confusion Matrix

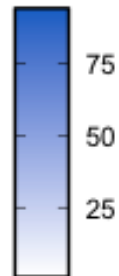


dip_y Removed

Probability Heatmap

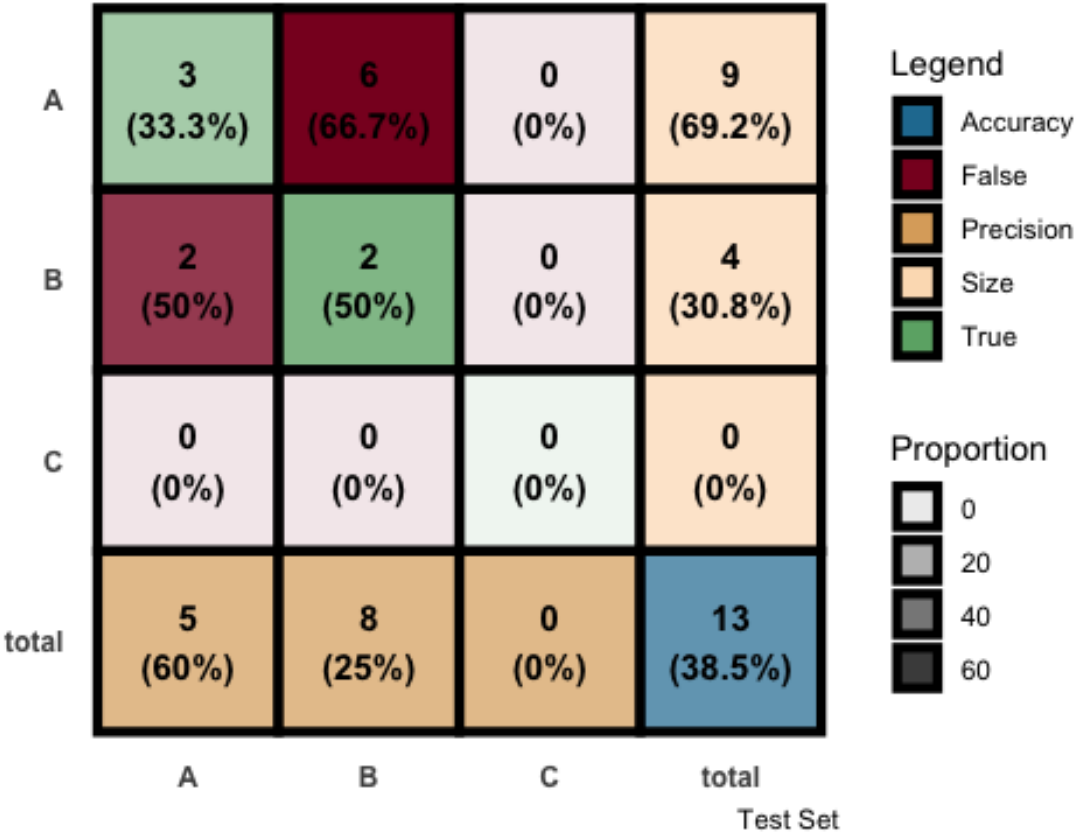
	A	B	C	Exp
2A	57	35	8	A
2G	49	40	11	A
2H	48	41	11	A
2L	43	44	13	A
2M	42	44	14	A
2O	45	42	12	A
2R	44	43	13	A
2S	23	49	28	A
2T	67	28	5	A
2U	46	42	12	A
2V	72	24	4	A
2J	40	45	15	A
2P	44	43	13	A
2Y	21	49	30	B
2Z	37	46	16	B
2AA	31	49	21	B
2AE	32	48	19	B
2AF	33	48	19	B
2AG	39	46	15	B
2AH	37	46	16	B
2AJ	33	48	19	B
2AK	34	48	18	B
2AM	22	49	29	B
2AO	39	46	16	B
2W	44	43	13	B
2X	40	45	15	B
2AB	24	49	27	B
2AC	32	48	20	B
2AP	39	46	15	B
2AQ	38	46	16	B
2AR	16	46	38	B
2AS	18	47	35	C
2AT	28	49	22	C
2AU	48	41	11	C
2AV	28	49	23	C
2AW	33	48	19	C
2AX	15	45	40	C
2AY	17	47	36	C
2AZ	22	49	29	C
2BA	1	6	94	C
2BB	1	9	89	C
2BC	7	33	59	C

% probability

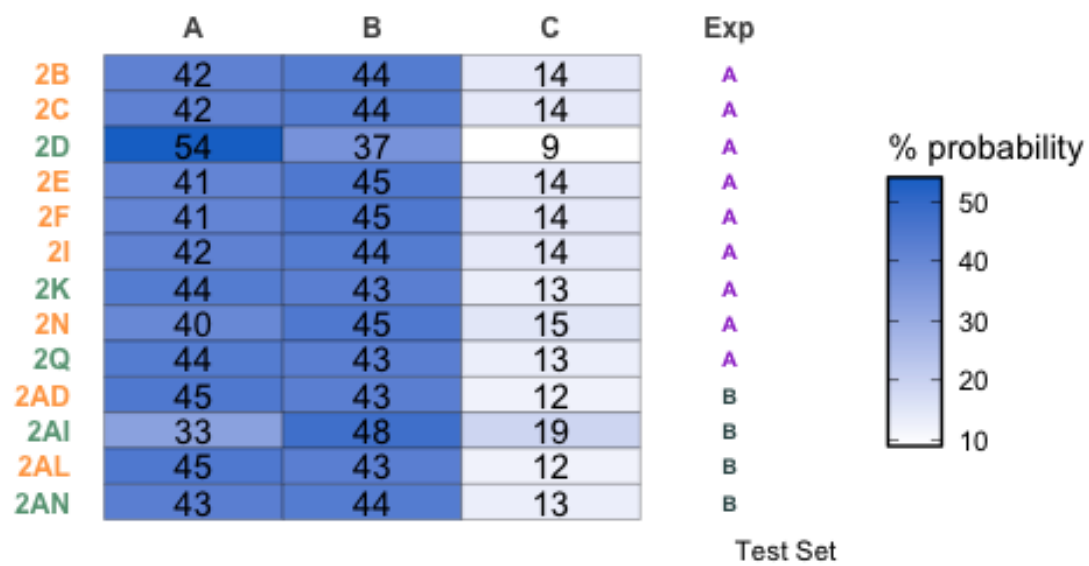


Training Set

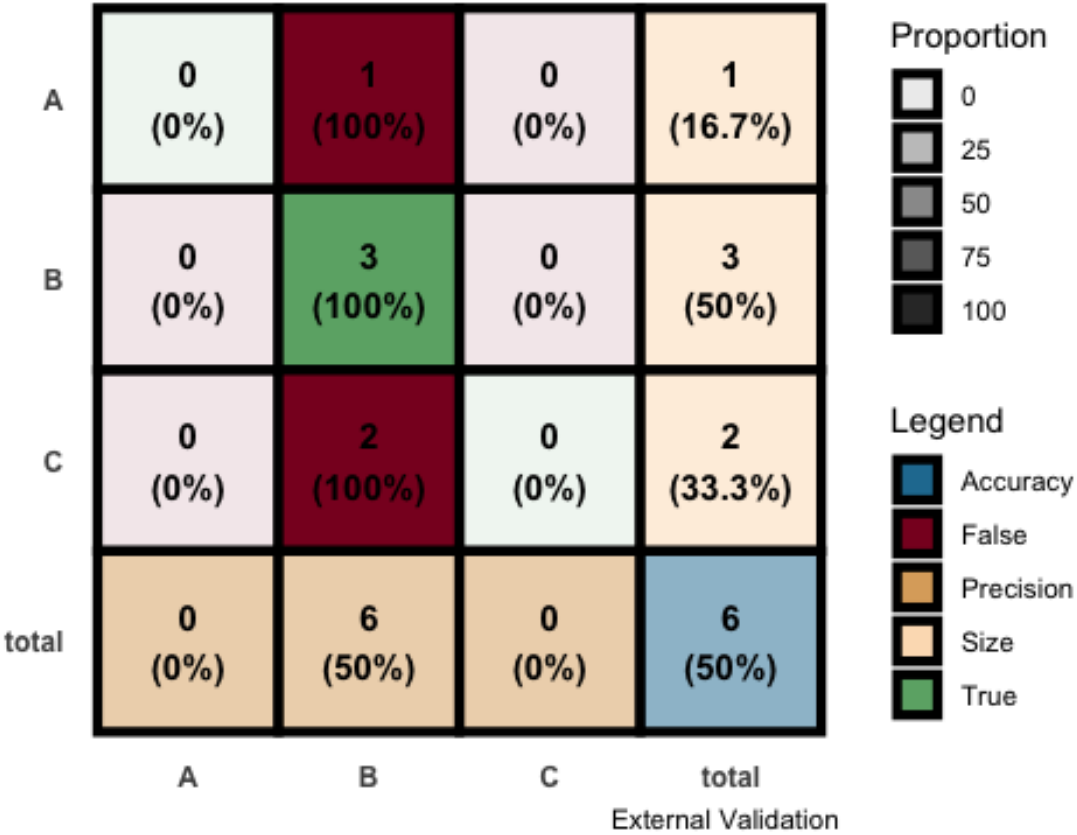
dip_y Removed
Confusion Matrix



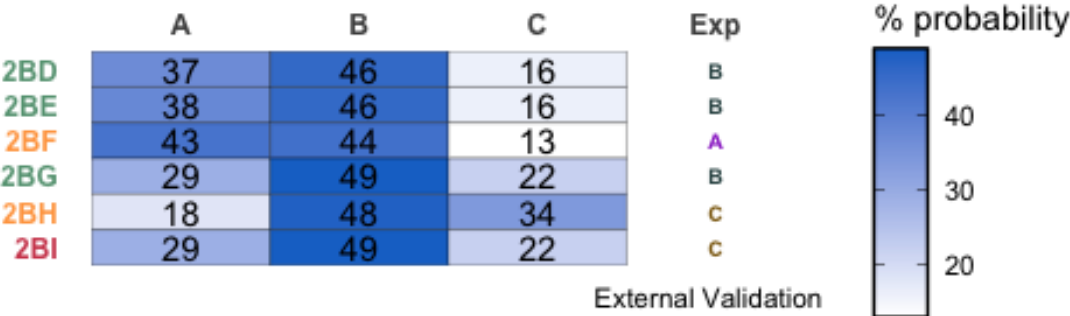
dip_y Removed Probability Heatmap



dip_y Removed
Confusion Matrix



dip_y Removed
Probability Heatmap



4. -2-3- Removed

formula

class ~ dip_y + Dist(2, 7) + NPA_1

Full Model Stats - Overall Accuracy and Pseudo-R2

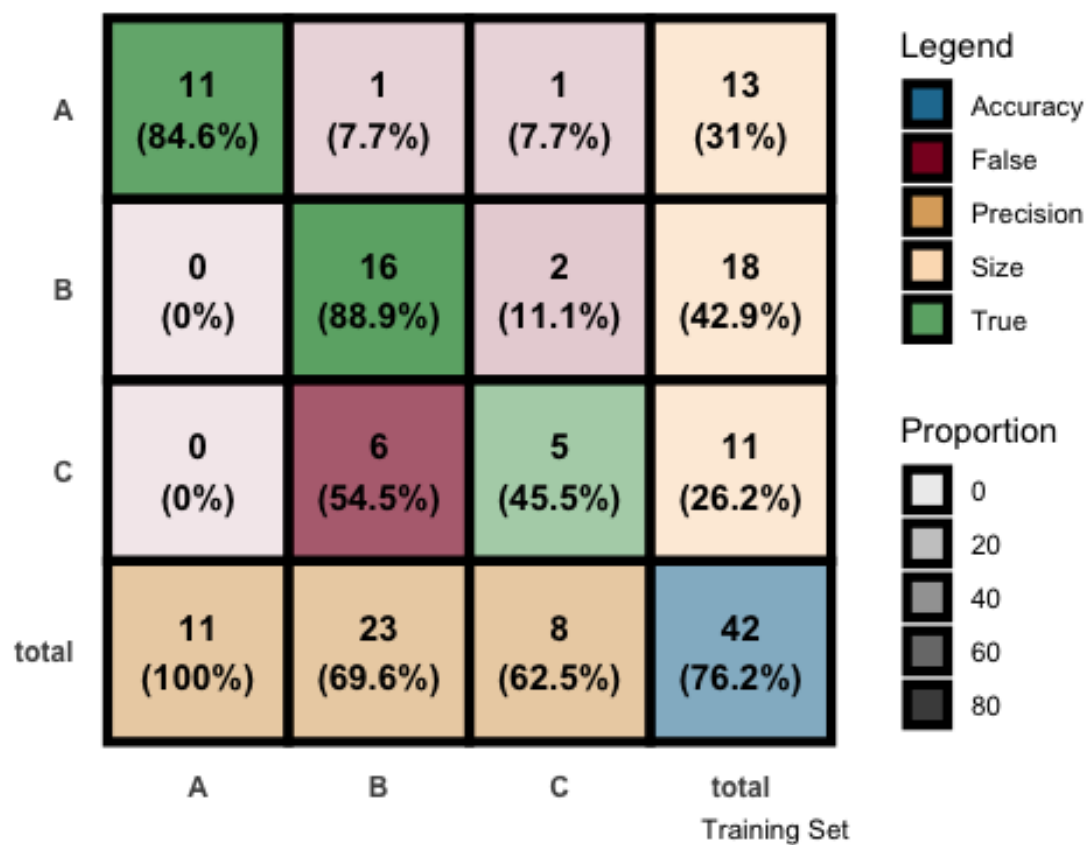
Accuracy	McFadden_R2
76.19%	0.344

Model Coefficients

	x
dip_y	-1.932190
Dist(2, 7)	-7.774941
NPA_1	23.524905

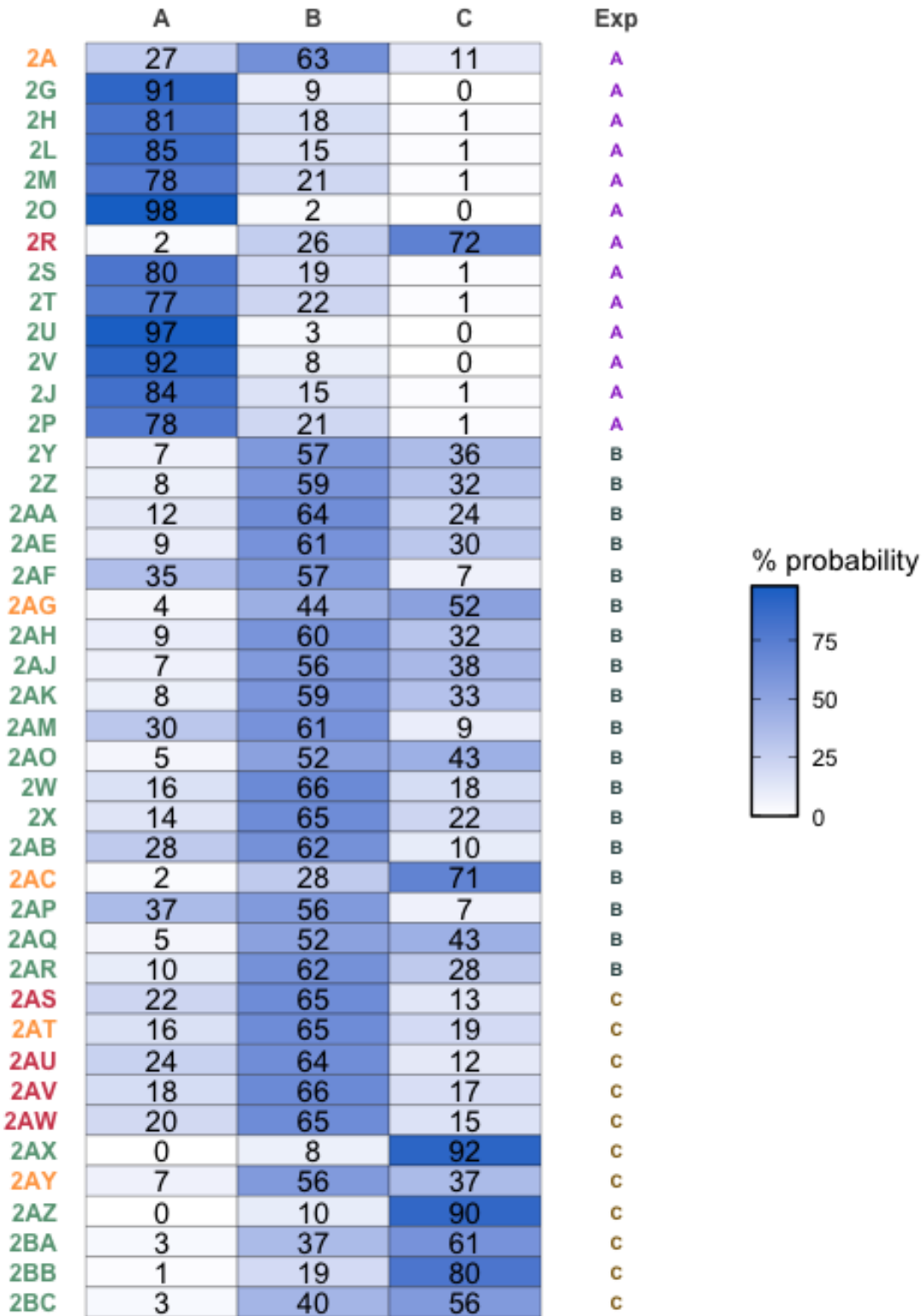
-2-3- Removed

Confusion Matrix



-2-3- Removed

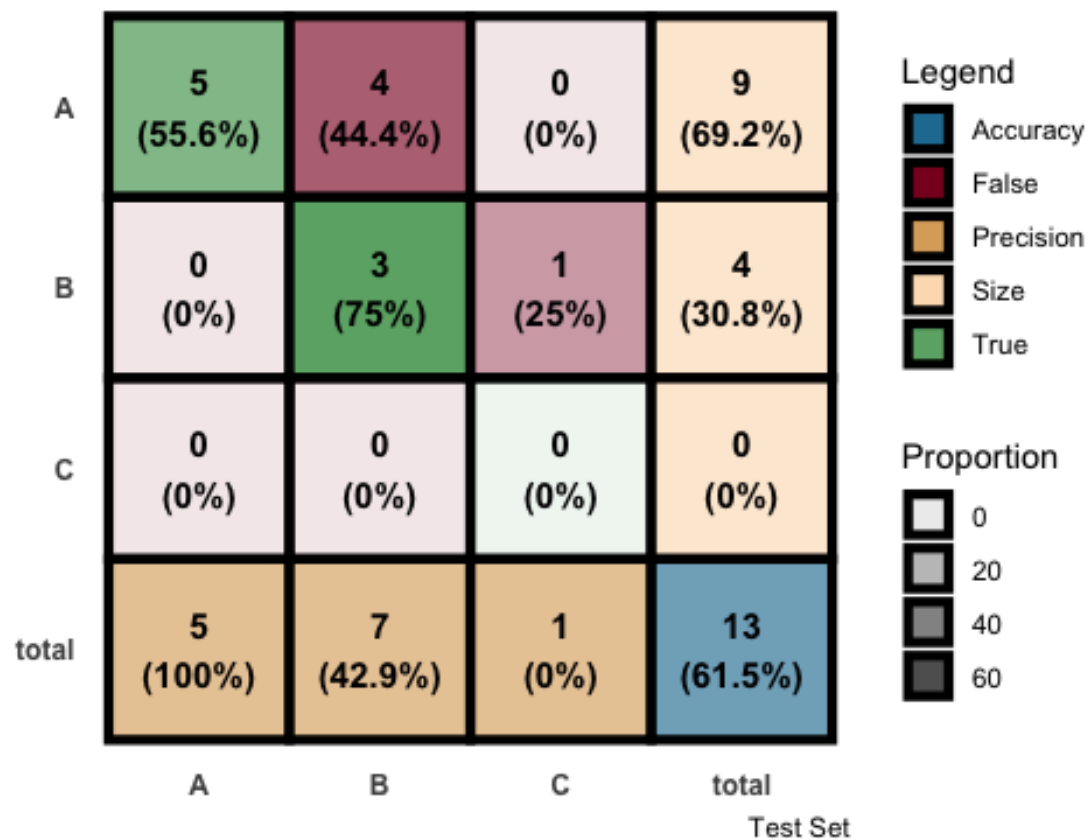
Probability Heatmap



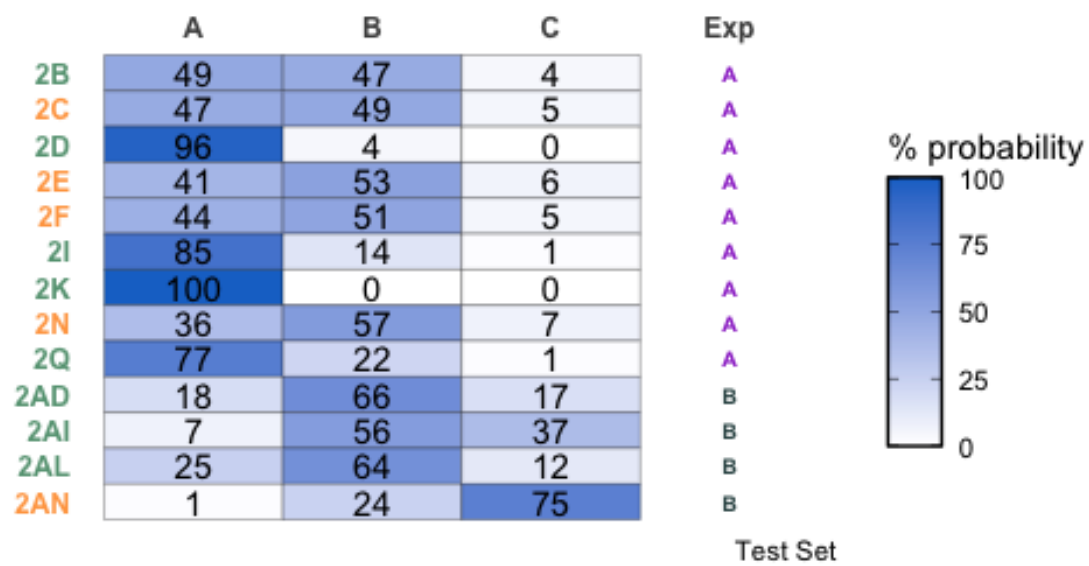
Training Set

-2-3- Removed

Confusion Matrix

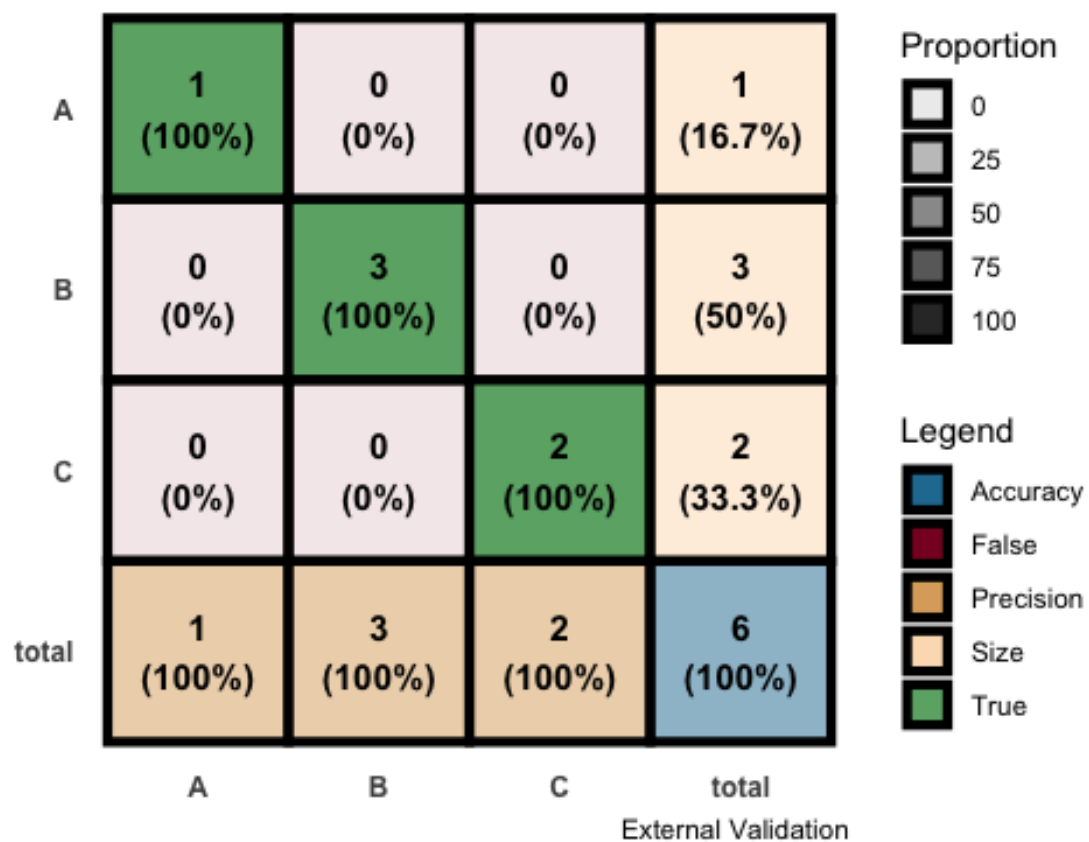


-2-3- Removed Probability Heatmap



-2-3- Removed

Confusion Matrix



-2-3- Removed

Probability Heatmap

