Aldehyde Deuteration - Best Model Feature Analysis

Table of Contents

Model Search	1
	1
Model Analysis - Variable Impac	t8
	8
3. dip_y Removed	22
	29
3. dip_y Removed	2

Model Search

1st Place

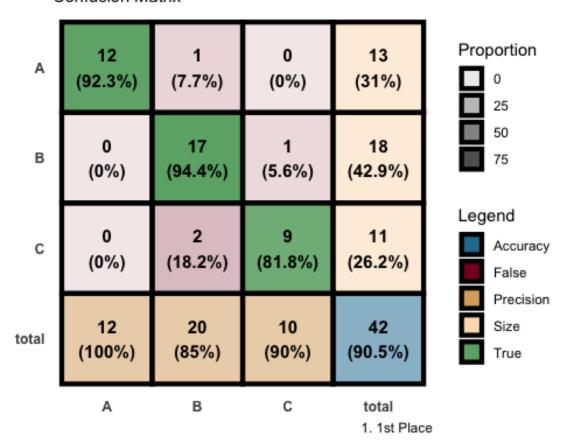
Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy McFadden_R2 90.48% 0.577

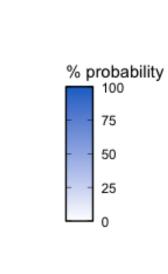
	X
-2-3-	-0.3801322
dip_y	-1.5680528
Dist(2, 7)	295.8494658
NPA_1	47.5647077

Training Set Confusion Matrix



Training Set Probability Heatmap

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



С

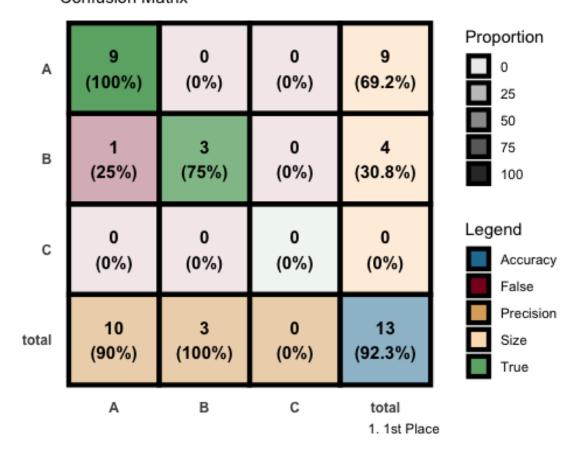
Exp

В

В

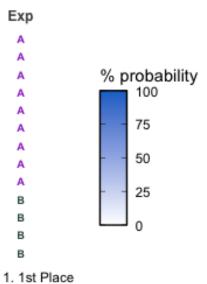
B B B B

Test Set Confusion Matrix

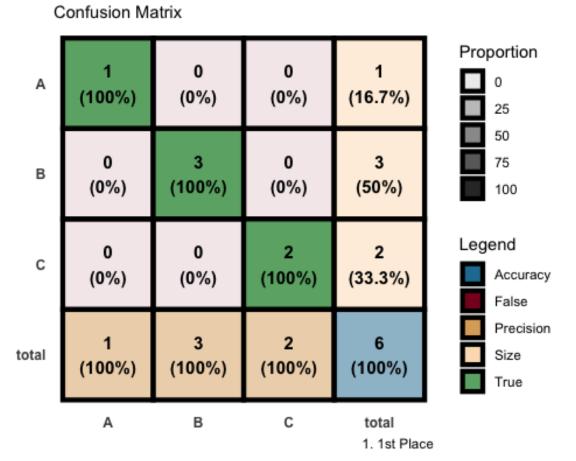


Test Set Probability Heatmap

	A	В	C
2B	86	14	0
2C	85	15	0
2D	100	0	0
2E	83	17	0
2F	81	19	0
21	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

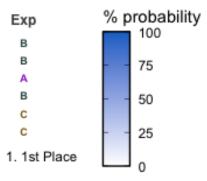


External Validation



External Validation Probability Heatmap

	Α	В	С
2BD	29	69	3
2BE	5	78	17
2BF	74	25	0
2BG	3	69	28
2BH	0	0	100
2BI	1	34	66



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
2B0	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

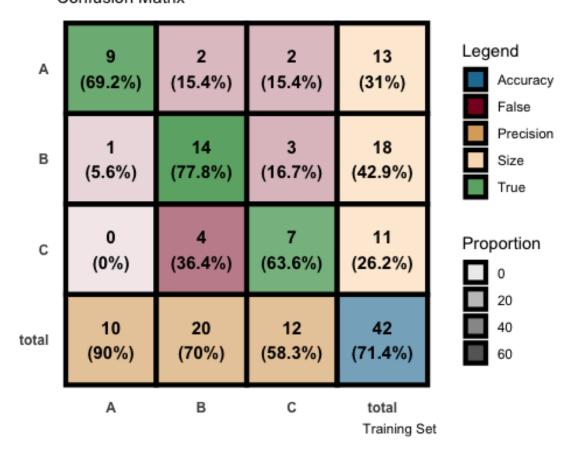
class
$$\sim$$
 -2-3- + dip_y + Dist(2, 7)

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
71.43%	0.338

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

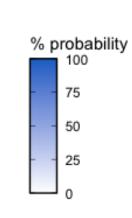
NPA_1 Removed Confusion Matrix



NPA_1 Removed

Probability	Heatmap
-------------	---------

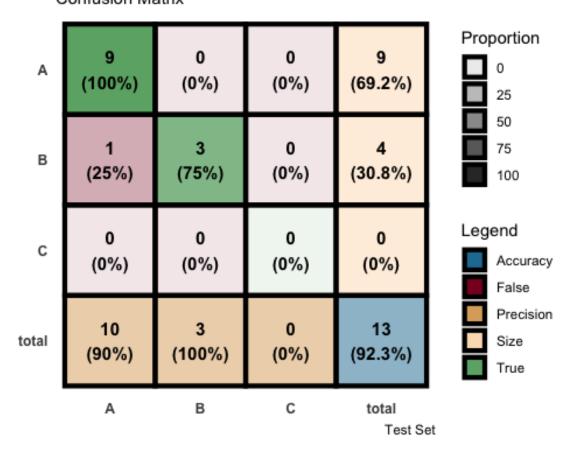
Α		В	С	
2A	16	65	19	
2G	90	10		
2H	89	10	1	
2L	66	31	2	
2M	75	31 23	1 1 2 1	
20	97	3	0	
2R		65	14	
2S	90	9	0	
2T	4	47	48	
2U	92	8	0	
2V	2	31	0 67	
2J	21 90 4 92 2 72 94	26	2	
2P	94	5	0	
2Y	27 3	63	11	
2Z	3	37	60	
2AA	3	36	61	
2AE	25	63	11	
2AF	61	36	3 8	
2AG	34	58	8	
2AH	26	63	11	
2AJ	4	46	50	
2AK	14	65	21 7 7	
2AM	35	57	7	
2AO	38	55	7	
2W	19	65	15	
2X	17	66	17 23	
2AB	17 12 29	64	23	
2AC	29	61	10	
2AP	15	65	20	
2AQ	38	56	7	
2AR	8 10	58 62	34	
2AS	10	62	28	
2AT	2	32	66	
2AU	20	65	15	
2AV	2	32	65	
2AW	8	58	35	
2AX	11	23	76	
2AY	7	56	37	
2AZ	0	10	90	
2BA	2 2 0	30	68	
2BB	2	34	63	
2BC	0	0	100	



С

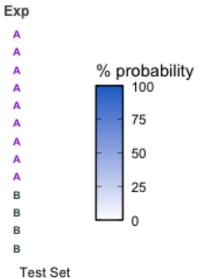
Exp

NPA_1 Removed Confusion Matrix

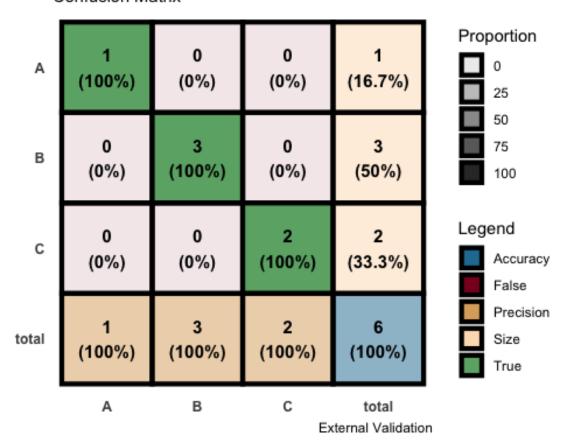


NPA_1 Removed Probability Heatmap

	Α	В	С
2B	79	20	1
2C	79	20	1
2D	97	3	0
2E	80	19	1
2F	75	24	1
21	78	21	1
2K	100	0	0
2N	75	23	1
2Q	94	6	0
2AD	19	65	15
2AI	12	64	23
2AL	54	43	4
2AN	18	65	16



NPA_1 Removed Confusion Matrix



NPA_1 Removed Probability Heatmap

	Α	В	С	Exp	% p	robability
2BD	36	57	7	В		
2BE	5	48	48	В		75
2BF	54	43	4	A		1.0
2BG	8	58	34	В		50
2BH	0	1	99	С		
2BI	4	44	52	С		25
			Ex	ternal Validation		0

2. Dist(2, 7) Removed

formula

$$class \sim -2-3- + dip_y + NPA_1$$

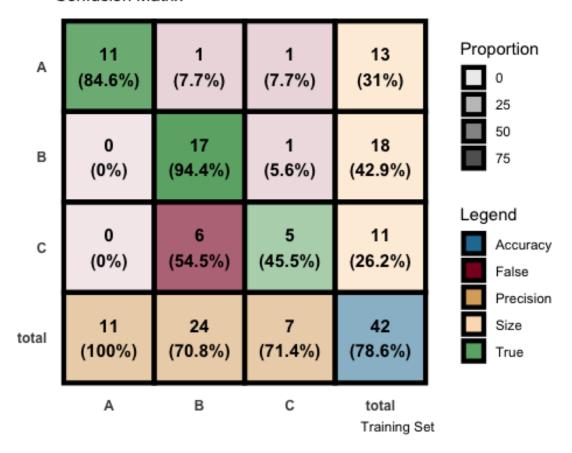
Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy McFadden_R2

78.57% 0.452

	X
-2-3-	-0.076705
dip_y	-1.830226
NPA_1	27.040639

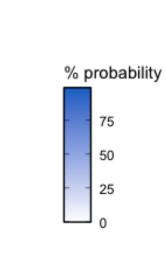
Dist(2, 7) Removed Confusion Matrix



Dist(2, 7) Removed

Probability Heatmap

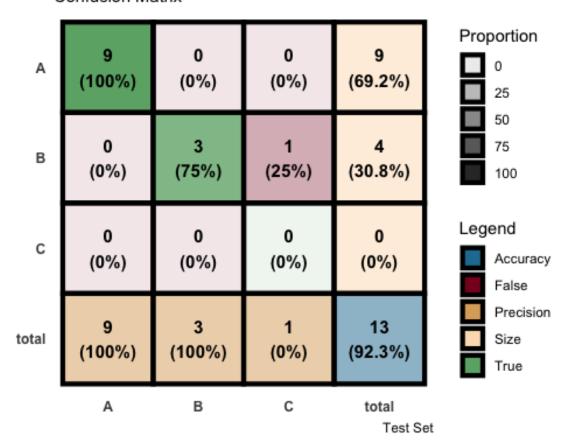
Α		В	С	
2A	29	63	7	
2G	94	5	0	
2H	89	10	0	
2L	88	12	0	
2M	84	16	1	
20	99	1	0	
2R	3	47	50	
2S	78	21	1	
2T	76	21 23	1	
2U	98	2	0	
2V	90	2 10	0	
2J	87	12	0	
2P	88	12 60	0	
2Y	6	60	34	
2Z	6	59	36	
2AA	7	62	31	
2AE	12	68	19	
2AF	45	51	4	
2AG	7	62	30	
2AH	13 7	68	19	
2AJ	7	61	32	
2AK	10	67	23 10	
2AM	23	67	10	
2AO	10	66	24	
2W	16	69	15	
2X	13	68	19	
2AB	20	68	12 61	
2AC	2	36	61	
2AP	34	60	6	
2AQ	10	66	24	
2AR	5 12	57 68	38	
2AS	12	68	19	
2AT	8	64	28	
2AU	24	66	10	
2AV	9	66	25	
2AW	19	68	13	
2AX	0	7	93	
2AY	4	49	47	
2AZ	0	10	90	
2BA	0	7	93	
2BB	0	8	92	
2BC	1	14	86	



Exp

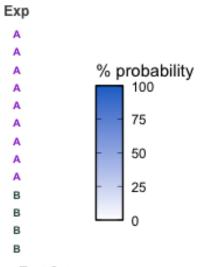
С

Dist(2, 7) Removed Confusion Matrix



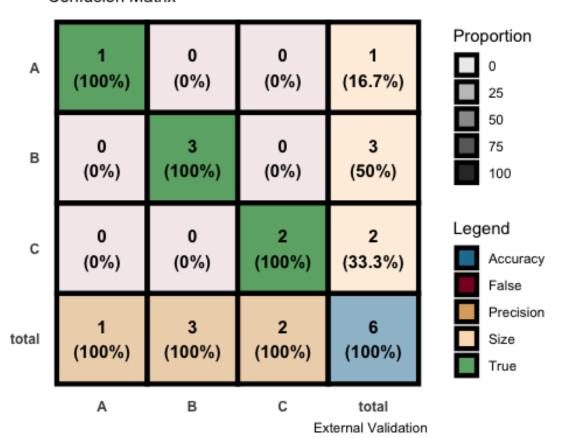
Dist(2, 7) Removed Probability Heatmap

	Α	В	С
2B	64	34	2
2C	62	34 36	2 2
2D	98	2	0
2E	57	41 39	3 2
2F 2I	58	39	2
21	89	11	0
2K	100	0	0
2N	51	46	3
2Q	87	12	1
2AD	18	69	14
2AI	8	65	27 5
2AL	38 3	65 57	5
2AN	3	43	54



Test Set

Dist(2, 7) Removed Confusion Matrix



Dist(2, 7) Removed Probability Heatmap

	Α	В	С	Exp	% prob	ability
2BD	24	66	10	В		
2BE	8	64	27	В	- 75	
2BF	68	31	2	A		
2BG	12	68	20	В	50	
2BH	0	3	97	С		
2BI	1	27	72	С	25	
External Validation						

3. dip_y Removed

formula

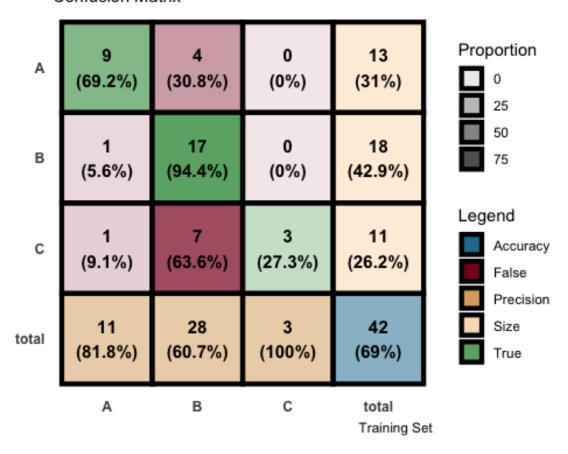
$$class \sim -2-3- + Dist(2, 7) + NPA_1$$

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy McFadden_R2 69.05% 0.217

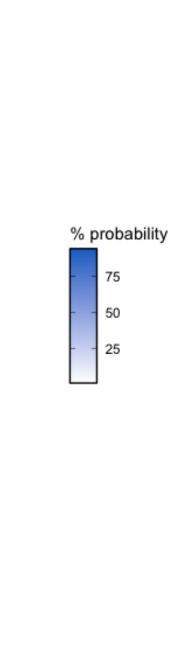
	X
-2-3-	-0.0879166
Dist(2, 7)	73.6506465
NPA_1	15.3035790

dip_y Removed Confusion Matrix



dip_y Removed Probability Heatmap

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



В

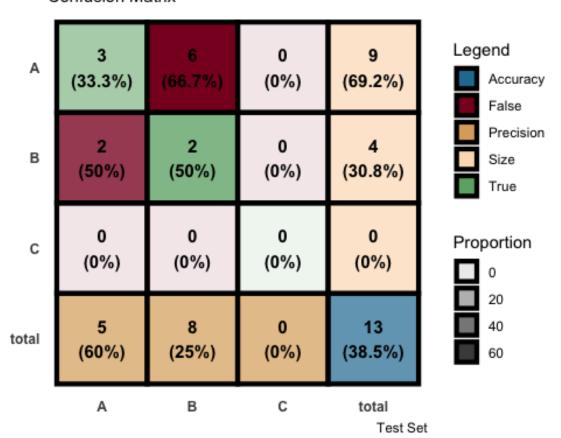
B B

B B

Exp

С

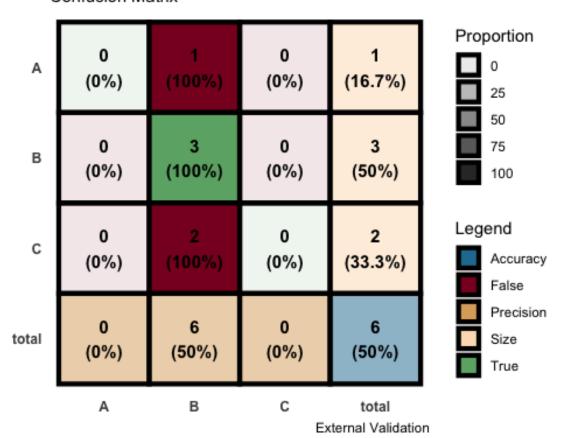
dip_y Removed Confusion Matrix



dip_y Removed Probability Heatmap

	Α	В	С	Exp	
2B	42	44	14	A	
2C	42	44	14	A	
2D	54	37	9	A	% probability
2E	41	45	14	A	
2F	41	45	14	A	50
21	42	44	14	A	40
2K	44	43	13	A	
2N	40	45	15	A	- 30
2Q	44	43	13	A	20
2AD	45	43	12	В	- 20
2AI	33	48	19	В	10
2AL	45	43	12	В	
2AN	43	44	13	В	
				Test Set	

dip_y Removed Confusion Matrix



dip_y Removed Probability Heatmap

	robubility i i	catinap			
	Α	В	С	Exp	% probability
2BD	37	46	16	В	
2BE	38	46	16	В	40
2BF	43	44	13	A	
2BG	29	49	22	В	30
2BH	18	48	34	С	
2BI	29	49	22	С	20
			Ex	ternal Validation	

4. -2-3- Removed

formula

class $\sim dip_y + Dist(2, 7) + NPA_1$

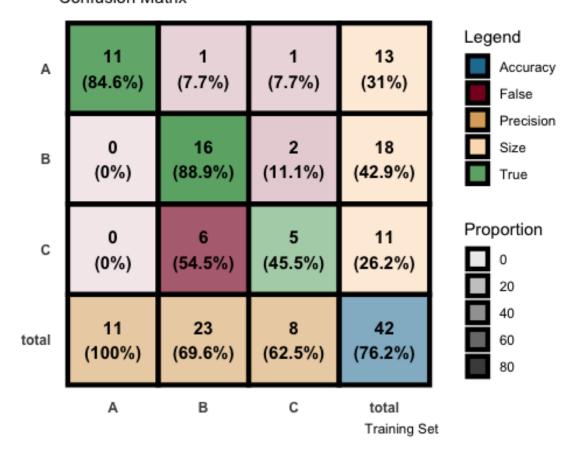
Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy McFadden_R2

76.19% 0.344

	X
dip_y	-1.932190
Dist(2, 7)	-7.774941
NPA 1	23.524905

-2-3- Removed Confusion Matrix



-2-3- Removed

Probability Heatmap

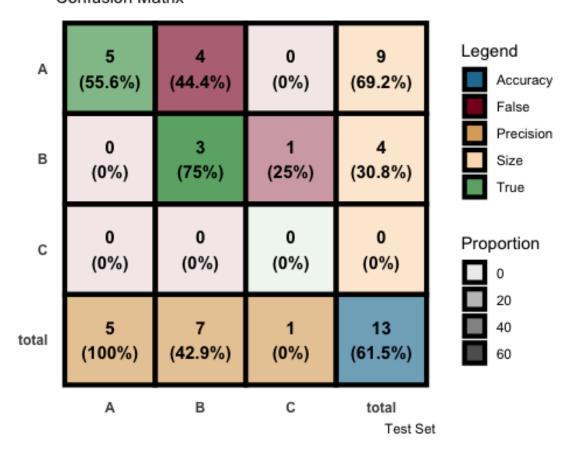
	Α	В	С
2A	27	63	11
2G	91	9	0
2H	81	18	1
2L	85	15	1
2M	78	21 2	1
20	98	2	0
2R	98 2 80	26 19	72
2S	80	19	1
2T	77 97	22	1
2U	97	3 8	0
2V	92	8	0
2J	84	15	1
2P	78	21	1
2Y	7	57	36
2Z	8	59	32
2AA	12	64	24
2AE	9	61	30
2AF	35	57	24 30 7
2AG	4	44	52
2AH	9 7	60	32
2AJ		56	38 33
2AK	8	59	33
2AM	30	61	9
2AO	5	52	43
2W	16	66	18 22 10
2X	14	65	22
2AB	28	62	10
2AC	2	28	71
2AP	37	56	7
2AQ	5 10	52 62	43
2AR	10	62	28
2AS	22	65	13
2AT	16	65	19
2AU	24	64	12
2AV	18	66	17
2AW	20	65	15
2AX	0	8	92
2AY	7	56	37
2AZ	0	10	90
2BA	3	37	61
2BB	1	19	80
2BC	3	40	56

% probability 75 50 25

Exp

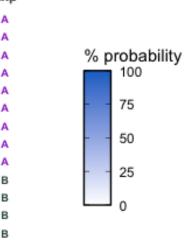
C C

-2-3- Removed Confusion Matrix



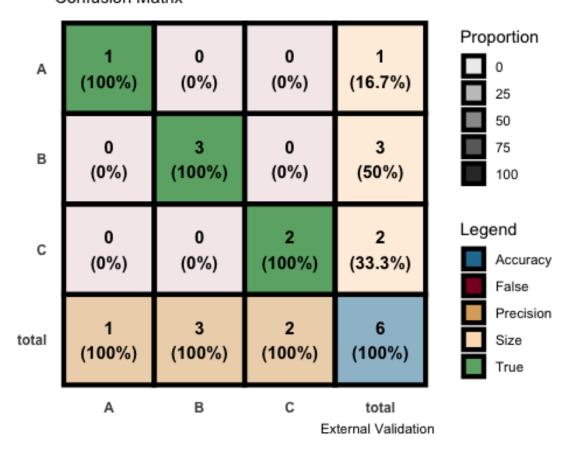
-2-3- Removed Probability Heatmap

	Α	В	С	Exp
2B	49	47	4	Α
2C	47	49	5	A
2D	96	4	0	A
2E	41	53	6	A
2F	44	51	5	A
21	85	14	1	Α
2K	100	0	0	Α
2N	36	57	7	Α
2Q	77	22	1	Α
2AD	18	66	17	В
2AI	7	56	37	В
2AL	25	64	12	В
2AN	1	24	75	В



Test Set

-2-3- Removed Confusion Matrix



-2-3- Removed Probability Heatmap

	Α	В	С	Exp	% p	robability
2BD	18	66	17	В		
2BE	20	65	14	В		75
2BF	59	38	3	A		
2BG	12	64	24	В		50
2BH	0	4	96	С		
2BI	1	20	79	С		25
			Ex	ternal Validation		,