1 Sub-Class Classification

The machine learning on the sub-classes was consistent with the main classes and on the surface appears superior, which is not 100% the case, as I will discuss below.

Table 1: **Summary of WSH Phone Call Data Table**

	# of		Class	Sub-Class Acc (%)	
Category	**Sub-Classes**	**Count**	**Accuracy**	**Train**	**Test**
APPOINTMENTS	5	45,340	85.1	90.7	85.8
ASK_A_DOCTOR	2	29,607	77.4	93.2	88.6
MISCELLANEOUS	5	34,384	72.0	93.0	79.1
LAB	3	14,293	87.5	95.6	86.1
PRESCRIPTION	5	48,804	72.3	93.2	88.5

Table 2: **APPOINTMENTS - Sub-Class Classification**

	Training						Test
Sub-Class	**Cases**	**Query**	**Late**	**Cancel**	**Reschedule**	**New**	**Total**
Query	2,381	116	2	10	24	310	462
Running Late	2,015	1	519	1	3	11	535
Cancellation	1,598	13	1	164	48	99	325
Reschedule	4,672	29	6	29	620	344	1,028
New Appt	34,674	116	8	39	105	5,868	6,136
Accuracy	90.7	**25.1**	**97.0**	**50.5**	**60.3**	**95.6**	85.9

The major issue here is the confusion between Query (QUERY ON CURRENT APPOINTMENT) and NEW APPOINTMENT. I would think that a couple of additional features specific to time references here would help resolve this issue.

Table 3: **ASK A DOCTOR - Sub-Class Classification**

	Training			Test
Sub-Class	**Cases**	**Medication**	**Symptoms**	**Total**
Medication Related	24,869	4143	165	4,308
Symptoms	4,738	410	143	553
Accuracy	93.2	**96.2**	**29.8**	88.6

This is a typical classification problem, where one class is unbalanced relative to another. As such, overall accuracy will appear to be quite high, while the smaller class will have quite poor accuracy, as we have here with only 29.8% for distinguishing when a Caller is describing symptoms. We might consider creating a third class here or at the very least figure out what is causing this confusion and how additional features may help.

Table 4: **MISCELLANEOUS - Sub-Class Classification**

	Training							Test
Sub-Class	**Cases**	**Sharing**	**Other**	**Provider**	**InsQuest**	**Pharmacy**	**Hospital**	**Total**
Sharing Information	9,010	1,427	516	62	5	.	6	2,016
Other	21,860	313	3,187	86	22	2	5	3,615
Change Provider	2,649	40	110	311	.	.	13	474
Insurance Question	364	8	37	.	15	.	.	60
Change Pharmacy	155	1	6	1	.	7	.	15
Change Hospital	346	7	16	53	.	.	8	84
Accuracy	79.1	**70.8**	**88.2**	**65.6**	**36.7**	**46.7**	**15.5**	79.3

I would definitely recommend consolidating this category and fold a couple of the categories together like Change Pharmacy and Change Hospital. Alternatively, the Other category maybe able to be broken into other classes.

Table 5: **LAB - Sub-Class Classification**

Sub-Class	Cases	Cancel	Results	Sharing	Total
Cancel	709	237	30	14	281
Results	10,015	6	2446	128	2,580
Sharing Information	3,569	8	365	720	1,093
Accuracy	87.5	**84.3**	**94.8**	**65.9**	86.1

This sub-class seems except that Sharing Information is showing up here, too. So, this sub-class should be rethought a little. Having it twice can make sense, but may be causing some overall confusion.

Table 6: **PRESCRIPTION - Sub-Class Classification**

Sub-Class	Cases	Follow-up	Pri-Auth	Provider	PharmQuest	Refill	Total
Follow-up	1,088	129	39	23	.	29	220
Prior Authorization	4,001	17	304	34	3	48	406
Provider	7.789	10	28	780	3	78	899
Queries from Pharmacy	4,796	.	10	6	460	40	516
Refill	31,130	5	17	69	8	1,940	2,039
Accuracy		**58.6**	**74.9**	**86.8**	**89.1**	**95.1**	88.6

This sub-class seems more or less fine, yet may have too many categories again, as most of the activity is in the Refill case.

2 Conclusion

The 21 sub-classes were designed with a lot of thought in mind and clearly contain linguistic features that distinguish them from one another. I would recommend narrowing a couple of the sub-classes and thinking a little more deeply about the Miscellaneous category, especially with respect to the Other sub-class. It may be worth running a 2-3 word topic models on each major to class to reconfirm the sub-class choices.