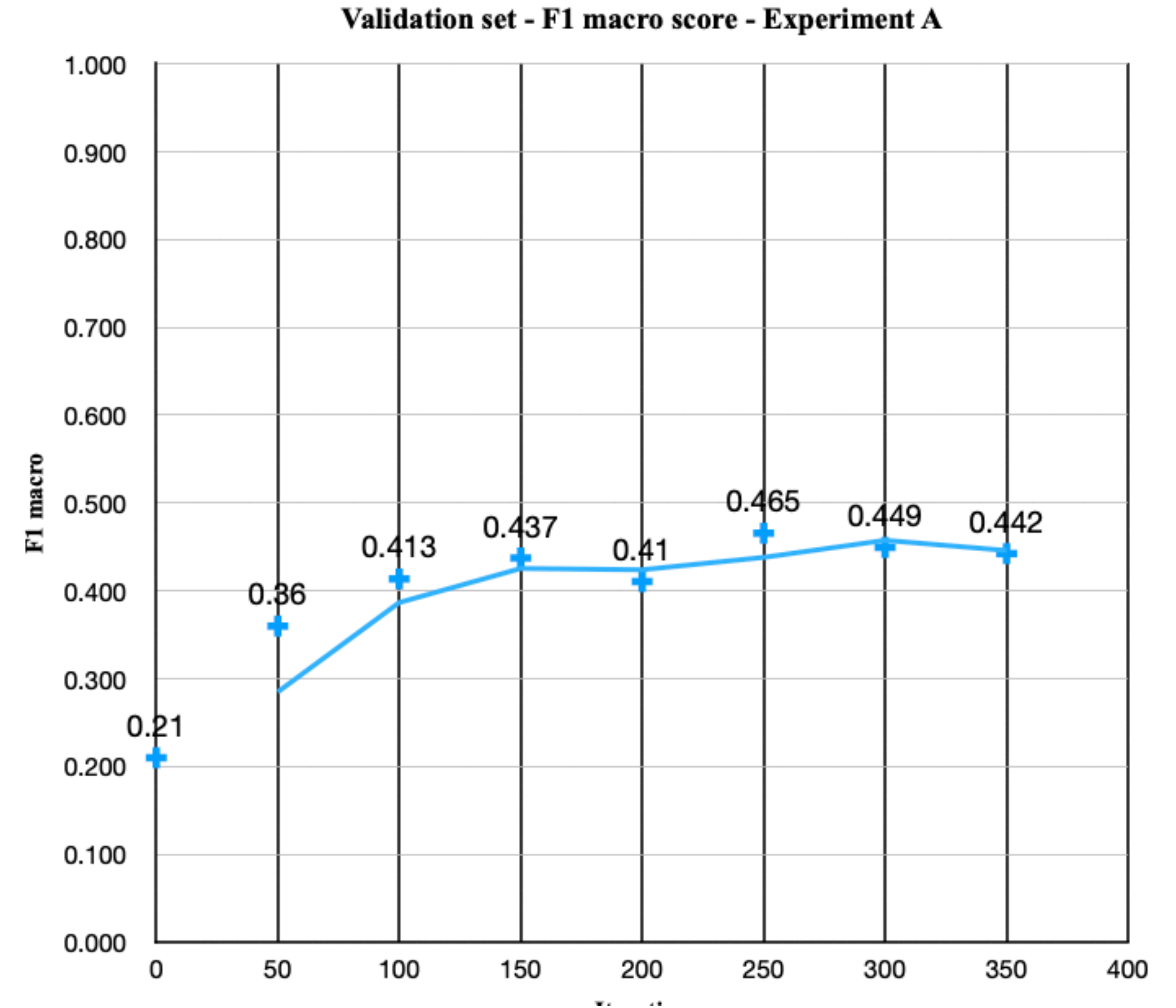


# Multi-label text classification

**Take home: DictateMD**

Srujan Vajram

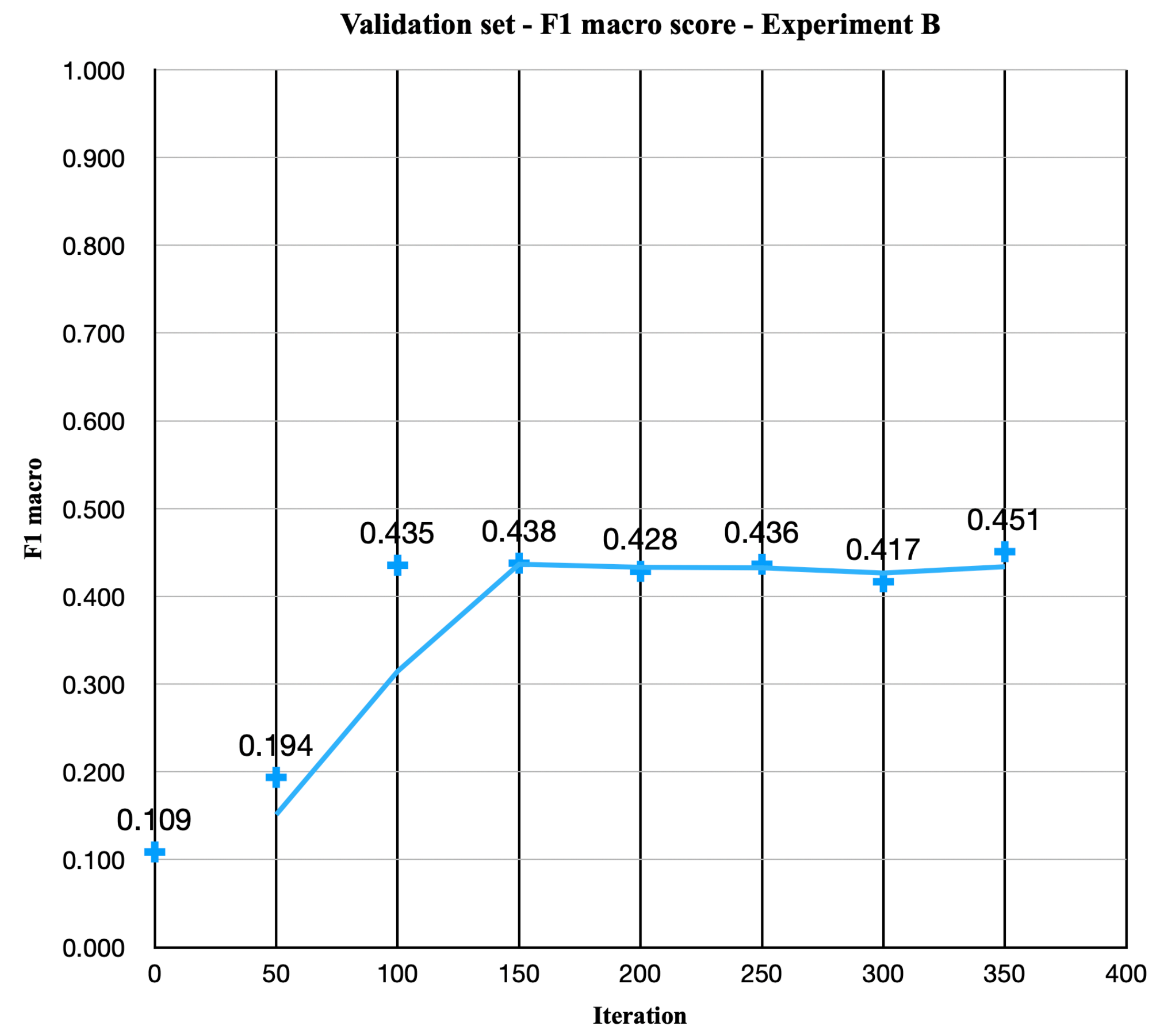
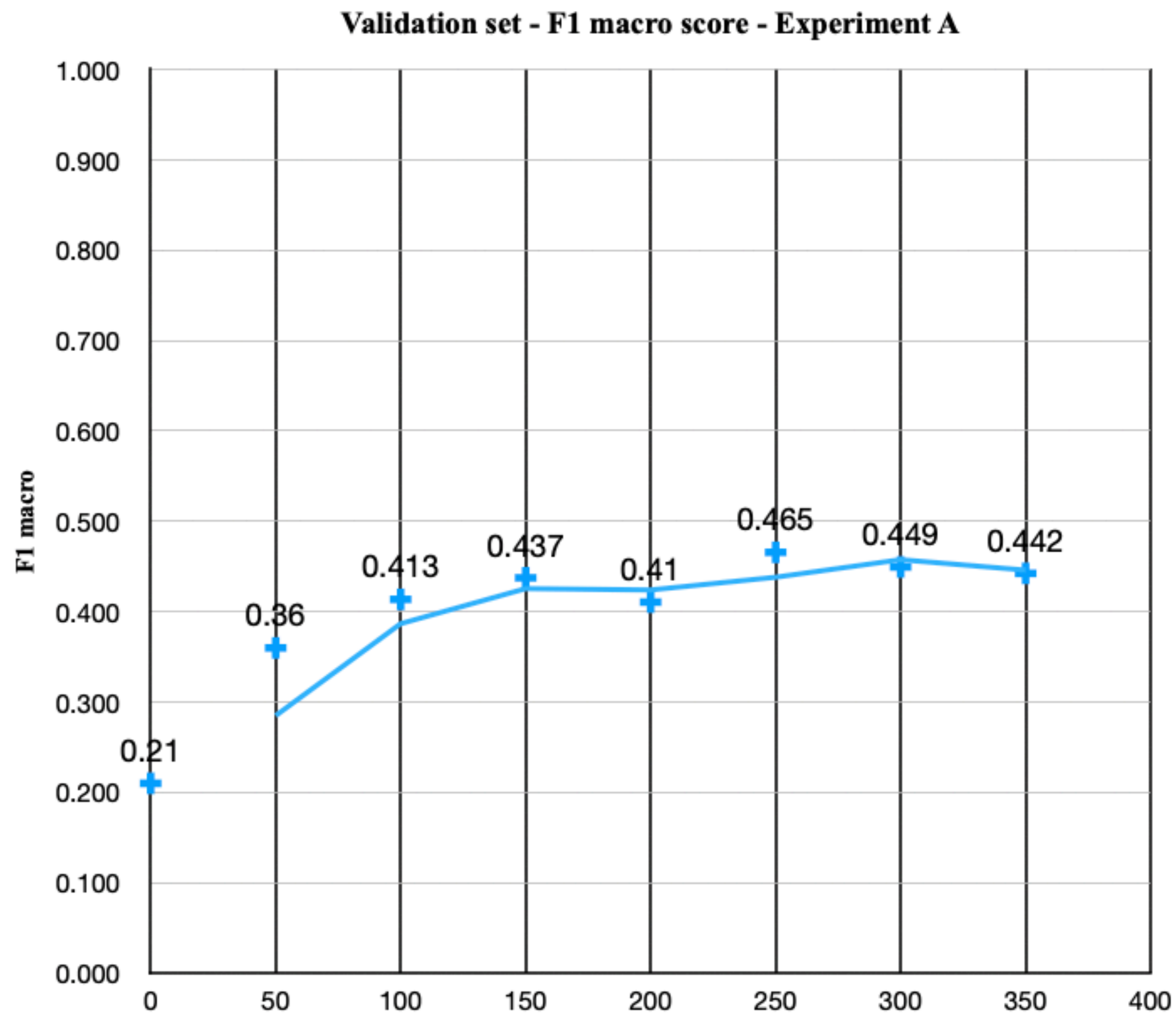
# Inspect training loss and macro f1 - Experiment A



# Precision, Recall, F1 - Experiment A

Label	Precision	Recall	F1	Support
Cardiovascular / Pulmonary	0.442	0.365	0.4	52
Surgery	0.453	0.358	0.4	162
Orthopedic	0.308	0.293	0.3	41
Obstetrics / Gynecology	0.574	0.625	0.598	56
Neurosurgery	0.286	0.286	0.286	14
Gastroenterology	0.2	0.083	0.118	12
Hematology - Oncology	0.435	0.833	0.571	24
Gastroenterology	0.459	0.739	0.567	23
ENT - Otolaryngology	0.4	0.333	0.364	12
Neurosurgery	0.692	0.643	0.667	14
Orthopedic	0.471	0.727	0.571	33
Neurology	0.571	0.308	0.4	13
Urology	0.528	0.559	0.543	34
Weighted average metrics	0.457	0.461	0.449	

# Comparing macro f1 - Experiment A vs B

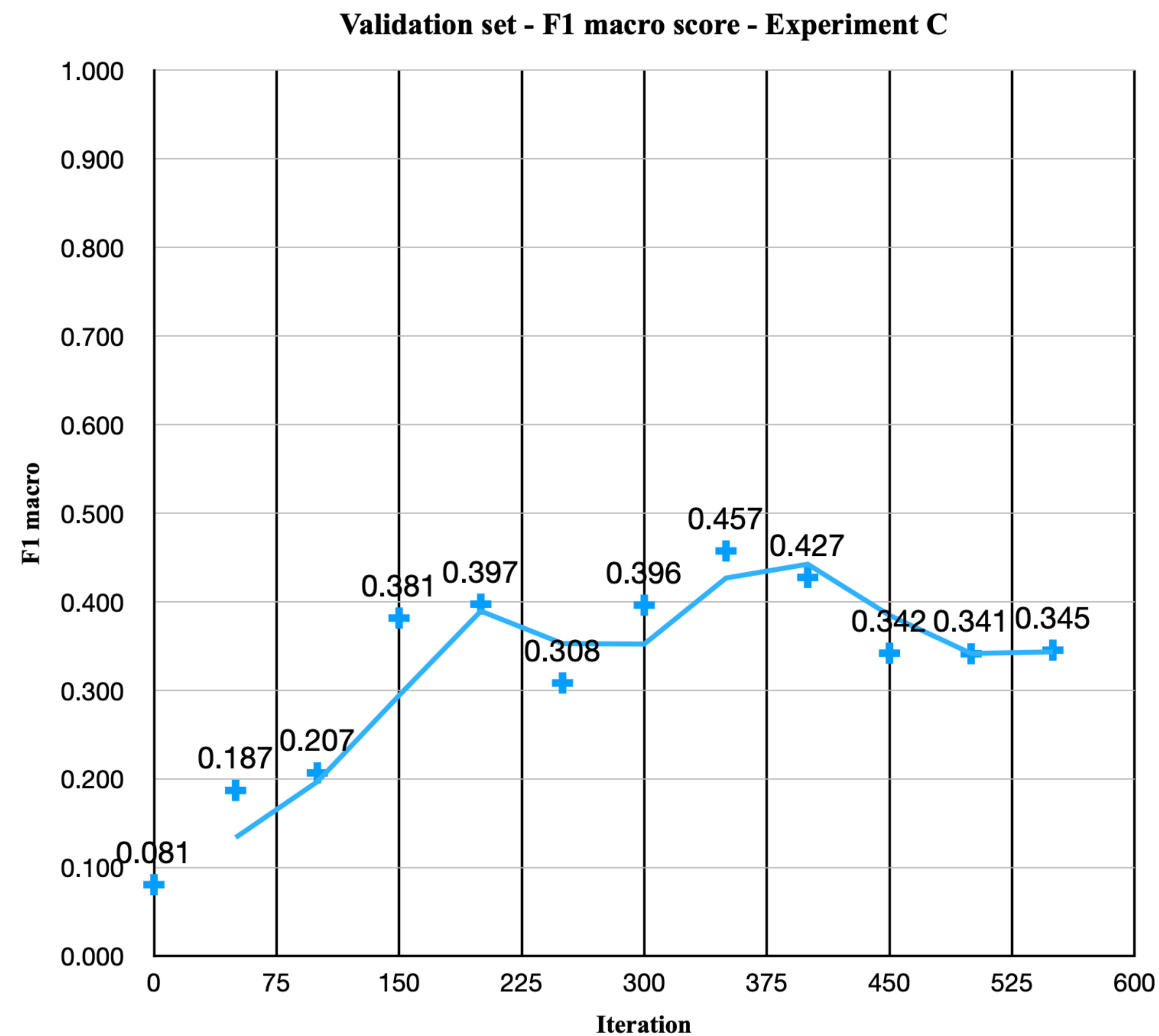
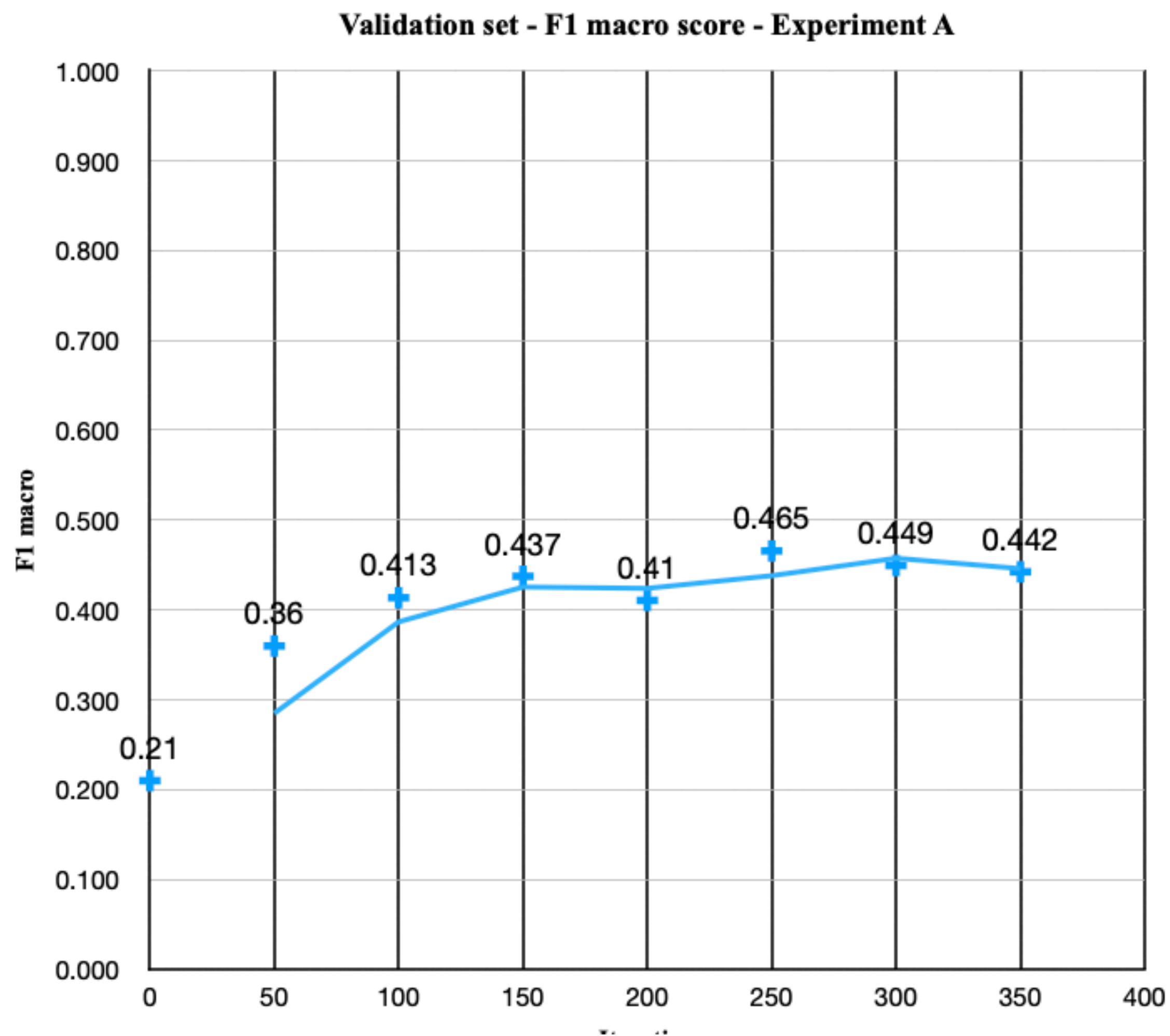




# Precision, Recall, F1 - Experiment B

Label	Precision	Recall	F1	Support
Cardiovascular / Pulmonary	0.568	0.618	0.592	52
Surgery	0.426	0.247	0.313	162
Orthopedic	0.5	0.333	0.4	41
Obstetrics / Gynecology	0.524	0.786	0.629	56
Neurosurgery	0.286	0.308	0.296	14
Gastroenterology	0.54	0.482	0.509	12
Hematology - Oncology	0.083	0.071	0.077	24
Gastroenterology	0.382	0.317	0.347	23
ENT - Otolaryngology	0.409	0.783	0.537	12
Neurosurgery	0.468	0.692	0.558	14
Orthopedic	0.55	0.917	0.688	33
Neurology	0.609	0.583	0.596	13
Urology	0.429	0.727	0.539	34
<b>Weighted average metrics</b>	<b>0.452</b>	<b>0.457</b>	<b>0.436</b>	

# Comparing macro f1 - Experiment A vs C





# Precision, Recall, F1 - Experiment C

Label	Precision	Recall	F1	Support
Cardiovascular / Pulmonary	0.443	0.659	0.529	52
Surgery	0.393	0.16	0.227	162
Orthopedic	0.373	0.5	0.427	41
Obstetrics / Gynecology	0.5	0.583	0.538	56
Neurosurgery	0.118	0.167	0.138	14
Gastroenterology	0.667	0.222	0.333	12
Hematology - Oncology	0.429	0.333	0.375	24
Gastroenterology	0.447	0.724	0.553	23
ENT - Otolaryngology	0.643	0.75	0.692	12
Neurosurgery	0.432	0.842	0.571	14
Orthopedic	0.27	0.27	0.27	33
Neurology	0.4	0.48	0.436	13
Urology	0.438	0.7	0.538	34
<b>Weighted average metrics</b>	<b>0.402</b>	<b>0.403</b>	<b>0.372</b>	

## Final test set Metrics

- Weighted Precision: 42.1%
- Weighted Recall: 41.0%
- Weighted F1 = 40.0%



# Future considerations

- We may need more data
- We may need to further consolidate labels (combine labels)
- Try a different pre-trained model