

Attention Mechanism Comparison: Simple vs Multi-Head

Performance Differences by Model and Code Smell

DataClass Smell

Model	Attention Type	F1	MCC	AUC	Precision	Recall	F1 Δ	MCC Δ
CodeBERT	Simple	0.6481	0.5935	0.8834	0.6036	0.7077	-0.0185	-0.0184
CodeBERT	Multi-Head	0.6296	0.5751	0.8807	0.5294	0.7788		
TokenIndexing	Simple	0.6257	0.5694	0.8495	0.6230	0.6356	-0.0014	-0.0036
TokenIndexing	Multi-Head	0.6243	0.5658	0.8700	0.5575	0.7156		
code2vec	Simple	0.5970	0.5363	0.8434	0.6010	0.5978	+0.0050	+0.0019
code2vec	Multi-Head	0.6020	0.5382	0.8677	0.5513	0.6684		
CuBERT	Simple	0.5909	0.5390	0.8775	0.4672	0.8100	-0.0012	-0.0096
CuBERT	Multi-Head	0.5897	0.5294	0.8553	0.4891	0.7487		

FeatureEnvy Smell

Model	Attention Type	F1	MCC	AUC	Precision	Recall	F1 Δ	MCC Δ
code2vec	Simple	0.3208	0.3206	0.8176	0.2455	0.5289	-0.0436	-0.0336
code2vec	Multi-Head	0.2772	0.2870	0.7837	0.2048	0.5889		
CuBERT	Simple	0.3120	0.3449	0.8724	0.2026	0.7145	+0.0290	+0.0083
CuBERT	Multi-Head	0.3410	0.3532	0.8169	0.2411	0.6200		
TokenIndexing	Simple	0.3063	0.3350	0.8579	0.2942	0.5673	+0.0506	+0.0326
TokenIndexing	Multi-Head	0.3569	0.3676	0.8874	0.2705	0.6109		
CodeBERT	Simple	0.2653	0.2993	0.8726	0.1649	0.6873	+0.0193	+0.0130
CodeBERT	Multi-Head	0.2846	0.3123	0.8576	0.1918	0.6673		

GodClass Smell

Model	Attention Type	F1	MCC	AUC	Precision	Recall	F1 Δ	MCC Δ
CodeBERT	Simple	0.5671	0.5161	0.9128	0.4718	0.7140	+0.0061	+0.0105
CodeBERT	Multi-Head	0.5732	0.5266	0.9095	0.4740	0.7417		
code2vec	Simple	0.5654	0.5227	0.9030	0.4454	0.7797	+0.0149	+0.0144
code2vec	Multi-Head	0.5803	0.5371	0.9009	0.4816	0.7521		
TokenIndexing	Simple	0.5544	0.5215	0.9158	0.4233	0.8297	-0.0020	-0.0186
TokenIndexing	Multi-Head	0.5524	0.5029	0.8878	0.4528	0.7200		
CuBERT	Simple	0.5177	0.4657	0.8841	0.4220	0.6872	+0.0035	+0.0004

Model	Attention Type	F1	MCC	AUC	Precision	Recall	F1 Δ	MCC Δ
CuBERT	Multi-Head	0.5212	0.4661	0.8780	0.4323	0.6652		

LongMethod Smell

Model	Attention Type	F1	MCC	AUC	Precision	Recall	F1 Δ	MCC Δ
TokenIndexing	Simple	0.8055	0.7854	0.9776	0.7473	0.8816	+0.0018	+0.0016
TokenIndexing	Multi-Head	0.8073	0.7870	0.9770	0.7361	0.8966		
code2vec	Simple	0.8021	0.7808	0.9638	0.7402	0.8811	+0.0155	+0.0157
code2vec	Multi-Head	0.8176	0.7965	0.9634	0.7776	0.8656		
CodeBERT	Simple	0.7808	0.7550	0.9744	0.7448	0.8250	+0.0033	+0.0046
CodeBERT	Multi-Head	0.7841	0.7596	0.9729	0.7217	0.8607		
CuBERT	Simple	0.7659	0.7392	0.9709	0.6880	0.8695	+0.0053	+0.0044
CuBERT	Multi-Head	0.7712	0.7436	0.9712	0.6965	0.8646		

Summary Statistics

Average Performance Improvement (Multi-Head vs Simple)

Smell	Avg F1 Δ	Avg MCC Δ	Winner
DataClass	-0.0040	-0.0074	Simple
FeatureEnvy	+0.0138	+0.0051	Multi-Head
GodClass	+0.0056	+0.0017	Multi-Head
LongMethod	+0.0065	+0.0066	Multi-Head

Key Findings

- **Multi-Head Attention** performs better on 3 out of 4 code smells (FeatureEnvy, GodClass, LongMethod)
- **Simple Attention** performs slightly better on DataClass detection
- **LongMethod** shows the most consistent improvement with multi-head attention across all models
- **FeatureEnvy** shows the largest improvement with multi-head attention (+0.0138 avg F1)
- Improvements are generally modest, suggesting both mechanisms are comparably effective for code smell detection