

# deberta-ell

July 30, 2024

```
[1]: # Clone the repository
!git clone https://github.com/arnavs04/deberta-ell
```

```
Cloning into 'deberta-ell'...
remote: Enumerating objects: 144, done.
remote: Counting objects: 100% (144/144), done.
remote: Compressing objects: 100% (104/104), done.
remote: Total 144 (delta 66), reused 104 (delta 34), pack-reused 0
Receiving objects: 100% (144/144), 7.59 MiB | 16.34 MiB/s, done.
Resolving deltas: 100% (66/66), done.
```

```
[2]: # Change to the repository directory
%cd deberta-ell
```

```
/kaggle/working/deberta-ell
```

```
[3]: %%capture
!pip install iterative-stratification
!pip install transformers
!pip install ptflops
```

```
[4]: !python src/train.py
```

```
Imports Done
tokenizer_config.json: 100%|| 52.0/52.0 [00:00<00:00, 342kB/s]
config.json: 100%|| 579/579 [00:00<00:00, 3.87MB/s]
spm.model: 100%|| 2.46M/2.46M [00:00<00:00, 3.68MB/s]
100%|| 3911/3911 [00:04<00:00, 884.34it/s]
----- Fold: 0 Training -----
----- Fold: 0 Training -----
----- Fold: 0 Training -----
DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",
  "hidden_dropout": 0.0,
  "hidden_dropout_prob": 0.0,
  "hidden_size": 768,
```

```

"initializer_range": 0.02,
"intermediate_size": 3072,
"layer_norm_eps": 1e-07,
"max_position_embeddings": 512,
"max_relative_positions": -1,
"model_type": "deberta-v2",
"norm_rel_ebd": "layer_norm",
"num_attention_heads": 12,
"num_hidden_layers": 12,
"output_hidden_states": true,
"pad_token_id": 0,
"pooler_dropout": 0,
"pooler_hidden_act": "gelu",
"pooler_hidden_size": 768,
"pos_att_type": [
    "p2c",
    "c2p"
],
"position_biased_input": false,
"position_buckets": 256,
"relative_attention": true,
"share_att_key": true,
"transformers_version": "4.42.3",
"type_vocab_size": 0,
"vocab_size": 128100
}

```

```

DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",
  "hidden_dropout": 0.0,
  "hidden_dropout_prob": 0.0,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-07,
  "max_position_embeddings": 512,
  "max_relative_positions": -1,
  "model_type": "deberta-v2",
  "norm_rel_ebd": "layer_norm",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_hidden_states": true,
  "pad_token_id": 0,
  "pooler_dropout": 0,
  "pooler_hidden_act": "gelu",

```

```

"pooler_hidden_size": 768,
"pos_att_type": [
    "p2c",
    "c2p"
],
"position_biased_input": false,
"position_buckets": 256,
"relative_attention": true,
"share_att_key": true,
"transformers_version": "4.42.3",
"type_vocab_size": 0,
"vocab_size": 128100
}

```

```

DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",
  "hidden_dropout": 0.0,
  "hidden_dropout_prob": 0.0,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-07,
  "max_position_embeddings": 512,
  "max_relative_positions": -1,
  "model_type": "deberta-v2",
  "norm_rel_ebd": "layer_norm",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_hidden_states": true,
  "pad_token_id": 0,
  "pooler_dropout": 0,
  "pooler_hidden_act": "gelu",
  "pooler_hidden_size": 768,
  "pos_att_type": [
    "p2c",
    "c2p"
  ],
  "position_biased_input": false,
  "position_buckets": 256,
  "relative_attention": true,
  "share_att_key": true,
  "transformers_version": "4.42.3",
  "type_vocab_size": 0,
  "vocab_size": 128100
}

```

```

pytorch_model.bin: 100%|| 371M/371M [00:09<00:00, 40.4MB/s]
Epoch: [1][0/366] Elapsed 0m 3.1048007011413574s (remain 18m 53.25225591659546s)
Loss: 2.3492(2.3492) Grad: inf LR: 0.00002000
Epoch: [1][20/366] Elapsed 0m 38.11789894104004s (remain 10m
26.222625459943515s) Loss: 0.1988(1.4906) Grad: 75012.1094 LR: 0.00001999
Epoch: [1][40/366] Elapsed 1m 10.793618440628052s (remain 9m
21.168926663515094s) Loss: 0.1328(0.8578) Grad: 118755.3672 LR: 0.00001996
Epoch: [1][60/366] Elapsed 1m 46.53714728355408s (remain 8m 52.685736417770386s)
Loss: 0.2068(0.6481) Grad: 157825.9375 LR: 0.00001991
Epoch: [1][80/366] Elapsed 2m 23.178016185760498s (remain 8m 23.77450139434245s)
Loss: 0.1789(0.5313) Grad: 117346.5312 LR: 0.00001985
Epoch: [1][100/366] Elapsed 2m 54.7576220035553s (remain 7m 38.522473573684806s)
Loss: 0.0948(0.4556) Grad: 49947.6992 LR: 0.00001977
Epoch: [1][120/366] Elapsed 3m 27.735506057739258s (remain 7m
0.6214792078192204s) Loss: 0.1018(0.4016) Grad: 70376.3438 LR: 0.00001967
Epoch: [1][140/366] Elapsed 3m 58.405025005340576s (remain 6m
20.433550540437068s) Loss: 0.0725(0.3630) Grad: 35987.1562 LR: 0.00001955
Epoch: [1][160/366] Elapsed 4m 30.046488285064697s (remain 5m 43.8480130337781s)
Loss: 0.2374(0.3335) Grad: 196335.2031 LR: 0.00001941
Epoch: [1][180/366] Elapsed 5m 9.170594215393066s (remain 5m 16.00309353507032s)
Loss: 0.0984(0.3110) Grad: 51232.3242 LR: 0.00001926
Epoch: [1][200/366] Elapsed 5m 46.96005439758301s (remain 4m 44.81795510249344s)
Loss: 0.1293(0.2923) Grad: 62935.1055 LR: 0.00001909
Epoch: [1][220/366] Elapsed 6m 22.195414543151855s (remain 4m
10.761697324692364s) Loss: 0.0846(0.2759) Grad: 25448.6191 LR: 0.00001890
Epoch: [1][240/366] Elapsed 6m 58.7888457775116s (remain 3m 37.214131627340066s)
Loss: 0.1329(0.2620) Grad: 42692.7734 LR: 0.00001870
Epoch: [1][260/366] Elapsed 7m 31.081170320510864s (remain 3m
1.4694363358377132s) Loss: 0.1115(0.2527) Grad: 59957.3555 LR: 0.00001848
Epoch: [1][280/366] Elapsed 8m 2.046936273574829s (remain 2m 25.81490954894616s)
Loss: 0.1539(0.2433) Grad: 50522.1133 LR: 0.00001824
Epoch: [1][300/366] Elapsed 8m 34.080013036727905s (remain 1m
51.013956303612304s) Loss: 0.1107(0.2343) Grad: 37699.5273 LR: 0.00001799
Epoch: [1][320/366] Elapsed 9m 12.322221279144287s (remain 1m
17.428348777450196s) Loss: 0.1472(0.2270) Grad: 32446.6504 LR: 0.00001773
Epoch: [1][340/366] Elapsed 9m 42.65092992782593s (remain 0m 42.7163438363508s)
Loss: 0.1238(0.2202) Grad: 86476.3594 LR: 0.00001745
Epoch: [1][360/366] Elapsed 10m 17.170334339141846s (remain 0m
8.54806557256427s) Loss: 0.1493(0.2141) Grad: 47679.5977 LR: 0.00001715
Epoch: [1][365/366] Elapsed 10m 26.91321086883545s (remain 0m 0.0s) Loss:
0.1136(0.2128) Grad: 47425.0977 LR: 0.00001708
EVAL: [0/62] Elapsed 0m 0.9985504150390625s (remain 1m 0.9115753173828125s)
Loss: 0.0799(0.0799)
EVAL: [20/62] Elapsed 0m 24.688971281051636s (remain 0m 48.202277263005584s)
Loss: 0.0749(0.1059)
EVAL: [40/62] Elapsed 0m 51.09206175804138s (remain 0m 26.16910480289924s) Loss:
0.0783(0.1056)

```

EVAL: [60/62] Elapsed 1m 13.010946273803711s (remain 0m 1.196900758586949s)  
 Loss: 0.0907(0.1072)  
 EVAL: [61/62] Elapsed 1m 13.048263788223267s (remain 0m 0.0s) Loss:  
 0.1109(0.1072)  
 Epoch 1 - avg\_train\_loss: 0.2128 avg\_val\_loss: 0.1072 time: 700s  
 Epoch 1 - avg\_train\_loss: 0.2128 avg\_val\_loss: 0.1072 time: 700s  
 Epoch 1 - avg\_train\_loss: 0.2128 avg\_val\_loss: 0.1072 time: 700s  
 Epoch 1 - Score: 0.4638 Scores: [0.5029585838714051, 0.44770457591027896,  
 0.42777679769707555, 0.47255310922944194, 0.47076414788668564,  
 0.46095217821770107]  
 Epoch 1 - Score: 0.4638 Scores: [0.5029585838714051, 0.44770457591027896,  
 0.42777679769707555, 0.47255310922944194, 0.47076414788668564,  
 0.46095217821770107]  
 Epoch 1 - Score: 0.4638 Scores: [0.5029585838714051, 0.44770457591027896,  
 0.42777679769707555, 0.47255310922944194, 0.47076414788668564,  
 0.46095217821770107]  
 Epoch 1 - Save Best Score: 0.4638 Model  
 Epoch 1 - Save Best Score: 0.4638 Model  
 Epoch 1 - Save Best Score: 0.4638 Model  
 Epoch: [2][0/366] Elapsed 0m 2.6623740196228027s (remain 16m  
 11.766517162322998s) Loss: 0.0839(0.0839) Grad: 177753.7344 LR: 0.00001706  
 Epoch: [2][20/366] Elapsed 0m 40.599266052246094s (remain 11m  
 6.987942286900079s) Loss: 0.0945(0.1016) Grad: 159408.2812 LR: 0.00001675  
 Epoch: [2][40/366] Elapsed 1m 19.40942144393921s (remain 10m  
 29.464926080005966s) Loss: 0.1468(0.1040) Grad: 315751.4688 LR: 0.00001643  
 Epoch: [2][60/366] Elapsed 1m 50.98014736175537s (remain 9m 14.900736808776855s)  
 Loss: 0.0748(0.1014) Grad: 200942.1094 LR: 0.00001610  
 Epoch: [2][80/366] Elapsed 2m 26.16205358505249s (remain 8m 34.273892243703244s)  
 Loss: 0.0736(0.0995) Grad: 158039.5781 LR: 0.00001575  
 Epoch: [2][100/366] Elapsed 3m 2.921147584915161s (remain 7m 59.94162485151014s)  
 Loss: 0.0969(0.1016) Grad: 229840.0469 LR: 0.00001540  
 Epoch: [2][120/366] Elapsed 3m 33.983524322509766s (remain 7m  
 13.272425281114806s) Loss: 0.1118(0.1012) Grad: 167960.0781 LR: 0.00001503  
 Epoch: [2][140/366] Elapsed 4m 6.11733865737915s (remain 6m 32.740434027732704s)  
 Loss: 0.0849(0.1005) Grad: 225059.4688 LR: 0.00001466  
 Epoch: [2][160/366] Elapsed 4m 40.98874616622925s (remain 5m  
 57.780701640229836s) Loss: 0.0934(0.1000) Grad: 174642.8281 LR: 0.00001427  
 Epoch: [2][180/366] Elapsed 5m 17.96368098258972s (remain 5m  
 24.990502661763003s) Loss: 0.0828(0.1000) Grad: 193098.4219 LR: 0.00001388  
 Epoch: [2][200/366] Elapsed 5m 47.8727707862854s (remain 4m 45.567199899189404s)  
 Loss: 0.1473(0.1003) Grad: 122951.8984 LR: 0.00001348  
 Epoch: [2][220/366] Elapsed 6m 23.448375463485718s (remain 4m  
 11.583775756585624s) Loss: 0.0967(0.0995) Grad: 62477.6992 LR: 0.00001308  
 Epoch: [2][240/366] Elapsed 6m 53.534637451171875s (remain 3m  
 34.48891983981946s) Loss: 0.0757(0.0989) Grad: 43012.2070 LR: 0.00001267  
 Epoch: [2][260/366] Elapsed 7m 26.921876192092896s (remain 2m  
 59.79615708877304s) Loss: 0.0640(0.0991) Grad: 111455.8672 LR: 0.00001225  
 Epoch: [2][280/366] Elapsed 8m 6.361610651016235s (remain 2m 27.12006016133944s)

Loss: 0.1269(0.1005) Grad: 211409.7344 LR: 0.00001183  
 Epoch: [2][300/366] Elapsed 8m 37.43734335899353s (remain 1m 51.73896119048027s)  
 Loss: 0.0900(0.1005) Grad: 160589.9531 LR: 0.00001141  
 Epoch: [2][320/366] Elapsed 9m 9.943146467208862s (remain 1m 17.094833616898427s) Loss: 0.1220(0.1009) Grad: 105277.4219 LR: 0.00001098  
 Epoch: [2][340/366] Elapsed 9m 43.18394994735718s (remain 0m 42.755421550392725s) Loss: 0.1540(0.1004) Grad: 122190.9297 LR: 0.00001056  
 Epoch: [2][360/366] Elapsed 10m 17.994941473007202s (remain 0m 8.559486724002909s) Loss: 0.1422(0.1011) Grad: 145574.8281 LR: 0.00001013  
 Epoch: [2][365/366] Elapsed 10m 25.236183643341064s (remain 0m 0.0s) Loss: 0.1276(0.1013) Grad: 63796.2344 LR: 0.00001002  
 EVAL: [0/62] Elapsed 0m 1.0379996299743652s (remain 1m 3.3179774284362793s) Loss: 0.0731(0.0731)  
 EVAL: [20/62] Elapsed 0m 24.25019907951355s (remain 0m 47.34562677428836s) Loss: 0.0772(0.1028)  
 EVAL: [40/62] Elapsed 0m 50.132888078689575s (remain 0m 25.67782072323125s) Loss: 0.0823(0.1038)  
 EVAL: [60/62] Elapsed 1m 12.05660605430603s (remain 0m 1.1812558369558417s) Loss: 0.0877(0.1051)  
 EVAL: [61/62] Elapsed 1m 12.093968868255615s (remain 0m 0.0s) Loss: 0.0889(0.1051)  
 Epoch 2 - avg\_train\_loss: 0.1013 avg\_val\_loss: 0.1051 time: 697s  
 Epoch 2 - avg\_train\_loss: 0.1013 avg\_val\_loss: 0.1051 time: 697s  
 Epoch 2 - avg\_train\_loss: 0.1013 avg\_val\_loss: 0.1051 time: 697s  
 Epoch 2 - Score: 0.4589 Scores: [0.49612453608035384, 0.44902431425939027, 0.42639326484782314, 0.474497072263456, 0.4639231527195707, 0.44328368761321335]  
 Epoch 2 - Score: 0.4589 Scores: [0.49612453608035384, 0.44902431425939027, 0.42639326484782314, 0.474497072263456, 0.4639231527195707, 0.44328368761321335]  
 Epoch 2 - Score: 0.4589 Scores: [0.49612453608035384, 0.44902431425939027, 0.42639326484782314, 0.474497072263456, 0.4639231527195707, 0.44328368761321335]  
 Epoch 2 - Save Best Score: 0.4589 Model  
 Epoch 2 - Save Best Score: 0.4589 Model  
 Epoch 2 - Save Best Score: 0.4589 Model  
 Epoch: [3][0/366] Elapsed 0m 1.4907596111297607s (remain 9m 4.127258062362671s) Loss: 0.0905(0.0905) Grad: 192476.7656 LR: 0.00001000  
 Epoch: [3][20/366] Elapsed 0m 37.47775650024414s (remain 10m 15.705999646868008s) Loss: 0.0794(0.0956) Grad: 259405.8281 LR: 0.00000957  
 Epoch: [3][40/366] Elapsed 1m 6.386371850967407s (remain 8m 46.233435404009924s) Loss: 0.0554(0.0950) Grad: 200251.1562 LR: 0.00000914  
 Epoch: [3][60/366] Elapsed 1m 39.75288367271423s (remain 8m 18.764418363571167s) Loss: 0.1022(0.0933) Grad: 359591.8125 LR: 0.00000872  
 Epoch: [3][80/366] Elapsed 2m 11.168203115463257s (remain 7m 41.51775170255587s) Loss: 0.0416(0.0935) Grad: 88321.1953 LR: 0.00000829  
 Epoch: [3][100/366] Elapsed 2m 46.9590322971344s (remain 7m 18.060827314263634s) Loss: 0.1247(0.0938) Grad: 175424.6406 LR: 0.00000787  
 Epoch: [3][120/366] Elapsed 3m 20.658919095993042s (remain 6m 46.292852715027266s) Loss: 0.0812(0.0930) Grad: 108741.9297 LR: 0.00000746  
 Epoch: [3][140/366] Elapsed 3m 59.834906578063965s (remain 6m

22.715276454357422s) Loss: 0.1019(0.0930) Grad: 207883.6875 LR: 0.00000704  
 Epoch: [3][160/366] Elapsed 4m 31.16875410079956s (remain 5m  
 45.276985035179564s) Loss: 0.1132(0.0930) Grad: 142288.0781 LR: 0.00000664  
 Epoch: [3][180/366] Elapsed 5m 1.7191205024719238s (remain 5m  
 8.386946369929888s) Loss: 0.1432(0.0936) Grad: 341878.9375 LR: 0.00000624  
 Epoch: [3][200/366] Elapsed 5m 37.23204970359802s (remain 4m 36.83227960743113s)  
 Loss: 0.1439(0.0957) Grad: 244966.6562 LR: 0.00000584  
 Epoch: [3][220/366] Elapsed 6m 11.939136028289795s (remain 4m 4.03246481494125s)  
 Loss: 0.0571(0.0941) Grad: 177393.4375 LR: 0.00000546  
 Epoch: [3][240/366] Elapsed 6m 46.98928737640381s (remain 3m  
 31.094028722201188s) Loss: 0.0941(0.0945) Grad: 107947.0391 LR: 0.00000508  
 Epoch: [3][260/366] Elapsed 7m 22.01667833328247s (remain 2m 57.82280162833206s)  
 Loss: 0.0911(0.0941) Grad: 104672.9141 LR: 0.00000471  
 Epoch: [3][280/366] Elapsed 7m 54.26590657234192s (remain 2m  
 23.461217290566083s) Loss: 0.0800(0.0940) Grad: 142112.6250 LR: 0.00000435  
 Epoch: [3][300/366] Elapsed 8m 32.43713617324829s (remain 1m 50.6591822301034s)  
 Loss: 0.0693(0.0936) Grad: 105397.5859 LR: 0.00000400  
 Epoch: [3][320/366] Elapsed 9m 7.960031747817993s (remain 1m  
 16.816826880535245s) Loss: 0.1016(0.0934) Grad: 279117.8750 LR: 0.00000367  
 Epoch: [3][340/366] Elapsed 9m 44.16873908042908s (remain 0m  
 42.827620167186865s) Loss: 0.1166(0.0935) Grad: 149843.6094 LR: 0.00000334  
 Epoch: [3][360/366] Elapsed 10m 15.613631010055542s (remain 0m  
 8.526504584626764s) Loss: 0.1229(0.0934) Grad: 237399.2500 LR: 0.00000303  
 Epoch: [3][365/366] Elapsed 10m 24.569028615951538s (remain 0m 0.0s) Loss:  
 0.0638(0.0932) Grad: 134360.4844 LR: 0.00000295  
 EVAL: [0/62] Elapsed 0m 1.0117652416229248s (remain 1m 1.717679738998413s) Loss:  
 0.0698(0.0698)  
 EVAL: [20/62] Elapsed 0m 24.223318576812744s (remain 0m 47.29314579282489s)  
 Loss: 0.0772(0.1007)  
 EVAL: [40/62] Elapsed 0m 50.11232948303223s (remain 0m 25.66729071082139s) Loss:  
 0.0751(0.1014)  
 EVAL: [60/62] Elapsed 1m 12.02895736694336s (remain 0m 1.1808025797859614s)  
 Loss: 0.0879(0.1026)  
 EVAL: [61/62] Elapsed 1m 12.066126823425293s (remain 0m 0.0s) Loss:  
 0.0776(0.1026)  
 Epoch 3 - avg\_train\_loss: 0.0932 avg\_val\_loss: 0.1026 time: 697s  
 Epoch 3 - avg\_train\_loss: 0.0932 avg\_val\_loss: 0.1026 time: 697s  
 Epoch 3 - avg\_train\_loss: 0.0932 avg\_val\_loss: 0.1026 time: 697s  
 Epoch 3 - Score: 0.4532 Scores: [0.4919739262336351, 0.44522549769490455,  
 0.41238511620272783, 0.4572670404762418, 0.4667607397140293, 0.4455408872645461]  
 Epoch 3 - Score: 0.4532 Scores: [0.4919739262336351, 0.44522549769490455,  
 0.41238511620272783, 0.4572670404762418, 0.4667607397140293, 0.4455408872645461]  
 Epoch 3 - Score: 0.4532 Scores: [0.4919739262336351, 0.44522549769490455,  
 0.41238511620272783, 0.4572670404762418, 0.4667607397140293, 0.4455408872645461]  
 Epoch 3 - Save Best Score: 0.4532 Model  
 Epoch 3 - Save Best Score: 0.4532 Model  
 Epoch 3 - Save Best Score: 0.4532 Model  
 Epoch: [4][0/366] Elapsed 0m 1.6589889526367188s (remain 10m 5.530967712402344s)

Loss: 0.0832(0.0832) Grad: 139194.5000 LR: 0.00000294  
 Epoch: [4][20/366] Elapsed 0m 34.970192432403564s (remain 9m 34.51030424662997s)  
 Loss: 0.0911(0.0896) Grad: 118955.6875 LR: 0.00000264  
 Epoch: [4][40/366] Elapsed 1m 13.504144191741943s (remain 9m 42.65480151990562s)  
 Loss: 0.0938(0.0901) Grad: 161525.8125 LR: 0.00000236  
 Epoch: [4][60/366] Elapsed 1m 43.52982711791992s (remain 8m 37.64913558959961s)  
 Loss: 0.1123(0.0907) Grad: 255816.0781 LR: 0.00000209  
 Epoch: [4][80/366] Elapsed 2m 17.263386726379395s (remain 8m 2.963768111334957s)  
 Loss: 0.0855(0.0894) Grad: 183216.2031 LR: 0.00000183  
 Epoch: [4][100/366] Elapsed 2m 48.708035945892334s (remain 7m 22.649797283776934s) Loss: 0.0873(0.0911) Grad: 195621.8438 LR: 0.00000159  
 Epoch: [4][120/366] Elapsed 3m 23.890732288360596s (remain 6m 52.83660669957317s) Loss: 0.1088(0.0926) Grad: 141838.0000 LR: 0.00000137  
 Epoch: [4][140/366] Elapsed 3m 57.94965124130249s (remain 6m 19.706890278674223s) Loss: 0.0776(0.0929) Grad: 153073.6719 LR: 0.00000116  
 Epoch: [4][160/366] Elapsed 4m 32.99287962913513s (remain 5m 47.59962934144539s)  
 Loss: 0.0771(0.0918) Grad: 247464.7812 LR: 0.00000097  
 Epoch: [4][180/366] Elapsed 5m 6.373036861419678s (remain 5m 13.143711709185823s) Loss: 0.0736(0.0904) Grad: 220666.4219 LR: 0.00000079  
 Epoch: [4][200/366] Elapsed 5m 36.82938861846924s (remain 4m 36.501736925609066s) Loss: 0.1005(0.0905) Grad: 135509.7031 LR: 0.00000063  
 Epoch: [4][220/366] Elapsed 6m 8.66800308227539s (remain 4m 1.8862463661987476s)  
 Loss: 0.0804(0.0907) Grad: 132213.0625 LR: 0.00000049  
 Epoch: [4][240/366] Elapsed 6m 42.729212522506714s (remain 3m 28.884446329101024s) Loss: 0.0721(0.0908) Grad: 140473.2500 LR: 0.00000037  
 Epoch: [4][260/366] Elapsed 7m 17.481844425201416s (remain 2m 55.99844315956375s) Loss: 0.0976(0.0904) Grad: 163770.6406 LR: 0.00000026  
 Epoch: [4][280/366] Elapsed 7m 51.54812049865723s (remain 2m 22.63911118286785s)  
 Loss: 0.0739(0.0900) Grad: 173022.7812 LR: 0.00000017  
 Epoch: [4][300/366] Elapsed 8m 27.955843210220337s (remain 1m 49.69146115835326s) Loss: 0.0629(0.0893) Grad: 94120.4922 LR: 0.00000010  
 Epoch: [4][320/366] Elapsed 9m 1.358715534210205s (remain 1m 15.891408719749165s) Loss: 0.0869(0.0893) Grad: 188874.7969 LR: 0.00000005  
 Epoch: [4][340/366] Elapsed 9m 40.35068702697754s (remain 0m 42.54770432749103s)  
 Loss: 0.0440(0.0890) Grad: 138987.6406 LR: 0.00000002  
 Epoch: [4][360/366] Elapsed 10m 17.799636602401733s (remain 0m 8.556781670393434s) Loss: 0.0881(0.0889) Grad: 160797.5156 LR: 0.00000000  
 Epoch: [4][365/366] Elapsed 10m 25.24880599975586s (remain 0m 0.0s) Loss: 0.0744(0.0887) Grad: 165842.7344 LR: 0.00000000  
 EVAL: [0/62] Elapsed 0m 1.0018727779388428s (remain 1m 1.1142394542694092s)  
 Loss: 0.0625(0.0625)  
 EVAL: [20/62] Elapsed 0m 24.21235704421997s (remain 0m 47.27174470538185s) Loss: 0.0739(0.0981)  
 EVAL: [40/62] Elapsed 0m 50.0992329120636s (remain 0m 25.660582711056975s) Loss: 0.0822(0.0996)  
 EVAL: [60/62] Elapsed 1m 12.018213748931885s (remain 0m 1.1806264549005192s)  
 Loss: 0.0886(0.1009)  
 EVAL: [61/62] Elapsed 1m 12.055418968200684s (remain 0m 0.0s) Loss:



```

0.0968(0.1008)
Epoch 4 - avg_train_loss: 0.0887 avg_val_loss: 0.1008 time: 697s
Epoch 4 - avg_train_loss: 0.0887 avg_val_loss: 0.1008 time: 697s
Epoch 4 - avg_train_loss: 0.0887 avg_val_loss: 0.1008 time: 697s
Epoch 4 - Score: 0.4493 Scores: [0.4852259149436871, 0.44405645506044156,
0.4081474032925275, 0.4530246787363071, 0.4642128928942826, 0.4412000026029325]
Epoch 4 - Score: 0.4493 Scores: [0.4852259149436871, 0.44405645506044156,
0.4081474032925275, 0.4530246787363071, 0.4642128928942826, 0.4412000026029325]
Epoch 4 - Score: 0.4493 Scores: [0.4852259149436871, 0.44405645506044156,
0.4081474032925275, 0.4530246787363071, 0.4642128928942826, 0.4412000026029325]
Epoch 4 - Save Best Score: 0.4493 Model
Epoch 4 - Save Best Score: 0.4493 Model
Epoch 4 - Save Best Score: 0.4493 Model
----- Fold: 0 Result -----
----- Fold: 0 Result -----
----- Fold: 0 Result -----
Score: 0.4493 Scores: [0.4852259149436871, 0.44405645506044156,
0.4081474032925275, 0.4530246787363071, 0.4642128928942826, 0.4412000026029325]
Score: 0.4493 Scores: [0.4852259149436871, 0.44405645506044156,
0.4081474032925275, 0.4530246787363071, 0.4642128928942826, 0.4412000026029325]
Score: 0.4493 Scores: [0.4852259149436871, 0.44405645506044156,
0.4081474032925275, 0.4530246787363071, 0.4642128928942826, 0.4412000026029325]
----- Fold: 1 Training -----
----- Fold: 1 Training -----
----- Fold: 1 Training -----
DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",
  "hidden_dropout": 0.0,
  "hidden_dropout_prob": 0.0,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-07,
  "max_position_embeddings": 512,
  "max_relative_positions": -1,
  "model_type": "deberta-v2",
  "norm_rel_ebd": "layer_norm",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_hidden_states": true,
  "pad_token_id": 0,
  "pooler_dropout": 0,
  "pooler_hidden_act": "gelu",
  "pooler_hidden_size": 768,
  "pos_att_type": [

```

```

        "p2c",
        "c2p"
    ],
    "position_biased_input": false,
    "position_buckets": 256,
    "relative_attention": true,
    "share_att_key": true,
    "transformers_version": "4.42.3",
    "type_vocab_size": 0,
    "vocab_size": 128100
}

```

```

DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",
  "hidden_dropout": 0.0,
  "hidden_dropout_prob": 0.0,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-07,
  "max_position_embeddings": 512,
  "max_relative_positions": -1,
  "model_type": "deberta-v2",
  "norm_rel_ebd": "layer_norm",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_hidden_states": true,
  "pad_token_id": 0,
  "pooler_dropout": 0,
  "pooler_hidden_act": "gelu",
  "pooler_hidden_size": 768,
  "pos_att_type": [
    "p2c",
    "c2p"
  ],
  "position_biased_input": false,
  "position_buckets": 256,
  "relative_attention": true,
  "share_att_key": true,
  "transformers_version": "4.42.3",
  "type_vocab_size": 0,
  "vocab_size": 128100
}

```

```

DebertaV2Config {

```

```

"name_or_path": "microsoft/deberta-v3-base",
"attention_dropout": 0.0,
"attention_probs_dropout_prob": 0.0,
"hidden_act": "gelu",
"hidden_dropout": 0.0,
"hidden_dropout_prob": 0.0,
"hidden_size": 768,
"initializer_range": 0.02,
"intermediate_size": 3072,
"layer_norm_eps": 1e-07,
"max_position_embeddings": 512,
"max_relative_positions": -1,
"model_type": "deberta-v2",
"norm_rel_ebd": "layer_norm",
"num_attention_heads": 12,
"num_hidden_layers": 12,
"output_hidden_states": true,
"pad_token_id": 0,
"pooler_dropout": 0,
"pooler_hidden_act": "gelu",
"pooler_hidden_size": 768,
"pos_att_type": [
  "p2c",
  "c2p"
],
"position_biased_input": false,
"position_buckets": 256,
"relative_attention": true,
"share_att_key": true,
"transformers_version": "4.42.3",
"type_vocab_size": 0,
"vocab_size": 128100
}

```

```

Epoch: [1][0/366] Elapsed 0m 3.385664224624634s (remain 20m 35.76744198799133s)
Loss: 2.4183(2.4183) Grad: inf LR: 0.00002000
Epoch: [1][20/366] Elapsed 0m 34.451833724975586s (remain 9m 25.9944111960275s)
Loss: 0.1549(1.4465) Grad: 101884.5312 LR: 0.00001999
Epoch: [1][40/366] Elapsed 1m 14.097216129302979s (remain 9m 47.35598151276747s)
Loss: 0.1182(0.8350) Grad: 87690.7109 LR: 0.00001996
Epoch: [1][60/366] Elapsed 1m 50.239503145217896s (remain 9m
11.197515726089478s) Loss: 0.2042(0.6223) Grad: 124227.9531 LR: 0.00001991
Epoch: [1][80/366] Elapsed 2m 27.77424430847168s (remain 8m 39.94641515943738s)
Loss: 0.1336(0.5017) Grad: 92776.0391 LR: 0.00001985
Epoch: [1][100/366] Elapsed 3m 0.7434895038604736s (remain 7m
54.22796751012902s) Loss: 0.1740(0.4259) Grad: 46497.5781 LR: 0.00001977
Epoch: [1][120/366] Elapsed 3m 35.33359932899475s (remain 7m
16.006048228129885s) Loss: 0.1289(0.3810) Grad: 105922.1641 LR: 0.00001967

```

Epoch: [1][140/366] Elapsed 4m 8.178016185760498s (remain 6m 36.02874923259651s)  
 Loss: 0.2106(0.3469) Grad: 63675.7188 LR: 0.00001955  
 Epoch: [1][160/366] Elapsed 4m 41.95574355125427s (remain 5m 59.011971602528774s) Loss: 0.0781(0.3194) Grad: 26287.8750 LR: 0.00001941  
 Epoch: [1][180/366] Elapsed 5m 16.399665117263794s (remain 5m 23.391922909910477s) Loss: 0.1171(0.2964) Grad: 42700.5117 LR: 0.00001926  
 Epoch: [1][200/366] Elapsed 5m 47.60997772216797s (remain 4m 45.35147424954084s) Loss: 0.0683(0.2795) Grad: 34381.2617 LR: 0.00001909  
 Epoch: [1][220/366] Elapsed 6m 17.440025806427002s (remain 4m 7.64164589109464s) Loss: 0.1134(0.2661) Grad: 76415.9844 LR: 0.00001890  
 Epoch: [1][240/366] Elapsed 6m 48.84120845794678s (remain 3m 32.05456870225464s) Loss: 0.0762(0.2540) Grad: 49763.1289 LR: 0.00001870  
 Epoch: [1][260/366] Elapsed 7m 23.687182664871216s (remain 2m 58.49484360081021s) Loss: 0.1478(0.2459) Grad: 42933.4219 LR: 0.00001848  
 Epoch: [1][280/366] Elapsed 7m 59.168317794799805s (remain 2m 24.944153069601384s) Loss: 0.0660(0.2362) Grad: 36240.5820 LR: 0.00001824  
 Epoch: [1][300/366] Elapsed 8m 37.74038505554199s (remain 1m 51.80440208840605s) Loss: 0.0982(0.2283) Grad: 69354.4141 LR: 0.00001799  
 Epoch: [1][320/366] Elapsed 9m 13.251182317733765s (remain 1m 17.558576960429946s) Loss: 0.1403(0.2225) Grad: 80001.5469 LR: 0.00001773  
 Epoch: [1][340/366] Elapsed 9m 42.624194622039795s (remain 0m 42.71438376994422s) Loss: 0.1411(0.2155) Grad: 39305.6250 LR: 0.00001745  
 Epoch: [1][360/366] Elapsed 10m 14.901786088943481s (remain 0m 8.516645236688987s) Loss: 0.2185(0.2103) Grad: 37311.4219 LR: 0.00001716  
 Epoch: [1][365/366] Elapsed 10m 25.858975887298584s (remain 0m 0.0s) Loss: 0.1159(0.2090) Grad: 55248.5000 LR: 0.00001708  
 EVAL: [0/62] Elapsed 0m 0.8663351535797119s (remain 0m 52.84644436836243s) Loss: 0.1413(0.1413)  
 EVAL: [20/62] Elapsed 0m 24.60849404335022s (remain 0m 48.045155037017096s) Loss: 0.1001(0.1360)  
 EVAL: [40/62] Elapsed 0m 47.506327390670776s (remain 0m 24.332509151319186s) Loss: 0.1336(0.1359)  
 EVAL: [60/62] Elapsed 1m 13.411676406860352s (remain 0m 1.2034701050304903s) Loss: 0.1517(0.1390)  
 EVAL: [61/62] Elapsed 1m 13.464263439178467s (remain 0m 0.0s) Loss: 0.0691(0.1389)  
 Epoch 1 - avg\_train\_loss: 0.2090 avg\_val\_loss: 0.1389 time: 699s  
 Epoch 1 - avg\_train\_loss: 0.2090 avg\_val\_loss: 0.1389 time: 699s  
 Epoch 1 - avg\_train\_loss: 0.2090 avg\_val\_loss: 0.1389 time: 699s  
 Epoch 1 - Score: 0.5293 Scores: [0.5583320874695651, 0.5419626480463531, 0.4390945624703074, 0.5373433817842864, 0.5583006998069013, 0.5406991070719418]  
 Epoch 1 - Score: 0.5293 Scores: [0.5583320874695651, 0.5419626480463531, 0.4390945624703074, 0.5373433817842864, 0.5583006998069013, 0.5406991070719418]  
 Epoch 1 - Score: 0.5293 Scores: [0.5583320874695651, 0.5419626480463531, 0.4390945624703074, 0.5373433817842864, 0.5583006998069013, 0.5406991070719418]  
 Epoch 1 - Save Best Score: 0.5293 Model  
 Epoch 1 - Save Best Score: 0.5293 Model  
 Epoch 1 - Save Best Score: 0.5293 Model

Epoch: [2][0/366] Elapsed 0m 1.5542232990264893s (remain 9m 27.29150414466858s)  
 Loss: 0.0809(0.0809) Grad: 225894.4531 LR: 0.00001707  
 Epoch: [2][20/366] Elapsed 0m 36.46808981895447s (remain 9m 59.11861845425199s)  
 Loss: 0.1005(0.1124) Grad: 185744.4219 LR: 0.00001676  
 Epoch: [2][40/366] Elapsed 1m 7.79793119430542s (remain 8m 57.42262532071368s)  
 Loss: 0.1288(0.1062) Grad: 56580.6484 LR: 0.00001644  
 Epoch: [2][60/366] Elapsed 1m 47.156004905700684s (remain 8m 55.78002452850342s)  
 Loss: 0.1267(0.1048) Grad: 96320.7891 LR: 0.00001610  
 Epoch: [2][80/366] Elapsed 2m 26.304396629333496s (remain 8m 34.77472888098828s)  
 Loss: 0.0994(0.1063) Grad: 87983.2500 LR: 0.00001576  
 Epoch: [2][100/366] Elapsed 2m 58.97048783302307s (remain 7m 49.57603243317942s)  
 Loss: 0.0786(0.1043) Grad: 123081.6094 LR: 0.00001540  
 Epoch: [2][120/366] Elapsed 3m 31.65705966949463s (remain 7m 8.561815033274343s)  
 Loss: 0.1135(0.1048) Grad: 93214.5391 LR: 0.00001504  
 Epoch: [2][140/366] Elapsed 3m 59.549455642700195s (remain 6m 22.25976964260667s)  
 Loss: 0.1196(0.1052) Grad: 114470.7266 LR: 0.00001466  
 Epoch: [2][160/366] Elapsed 4m 38.183512926101685s (remain 5m 54.208820806527s)  
 Loss: 0.0707(0.1069) Grad: 115921.0781 LR: 0.00001428  
 Epoch: [2][180/366] Elapsed 5m 11.397531747817993s (remain 5m 18.27924515660959s)  
 Loss: 0.1356(0.1072) Grad: 83828.7734 LR: 0.00001389  
 Epoch: [2][200/366] Elapsed 5m 48.522563219070435s (remain 4m 46.10061159774432s)  
 Loss: 0.1528(0.1071) Grad: 241205.1094 LR: 0.00001349  
 Epoch: [2][220/366] Elapsed 6m 21.886981964111328s (remain 4m 10.559332057901088s)  
 Loss: 0.1344(0.1059) Grad: 99006.1953 LR: 0.00001309  
 Epoch: [2][240/366] Elapsed 6m 58.30974268913269s (remain 3m 36.96563417486141s)  
 Loss: 0.0797(0.1055) Grad: 83784.0000 LR: 0.00001268  
 Epoch: [2][260/366] Elapsed 7m 34.16788363456726s (remain 3m 2.711217554136283s)  
 Loss: 0.0901(0.1053) Grad: 69655.6016 LR: 0.00001226  
 Epoch: [2][280/366] Elapsed 8m 9.638828754425049s (remain 2m 28.111389480875914s)  
 Loss: 0.0947(0.1050) Grad: 107263.2188 LR: 0.00001184  
 Epoch: [2][300/366] Elapsed 8m 42.92822790145874s (remain 1m 52.924700377391446s)  
 Loss: 0.1064(0.1043) Grad: 90821.8281 LR: 0.00001142  
 Epoch: [2][320/366] Elapsed 9m 16.430615186691284s (remain 1m 18.00429184860161s)  
 Loss: 0.1180(0.1038) Grad: 94608.3125 LR: 0.00001099  
 Epoch: [2][340/366] Elapsed 9m 52.57464051246643s (remain 0m 43.44388860062065s)  
 Loss: 0.0925(0.1036) Grad: 94352.3672 LR: 0.00001057  
 Epoch: [2][360/366] Elapsed 10m 24.780139446258545s (remain 0m 8.653464535266721s)  
 Loss: 0.0982(0.1028) Grad: 122683.5703 LR: 0.00001014  
 Epoch: [2][365/366] Elapsed 10m 32.75616669654846s (remain 0m 0.0s) Loss:  
 0.0970(0.1027) Grad: 104785.4766 LR: 0.00001003  
 EVAL: [0/62] Elapsed 0m 0.8740301132202148s (remain 0m 53.315836906433105s)  
 Loss: 0.1126(0.1126)  
 EVAL: [20/62] Elapsed 0m 23.95057988166809s (remain 0m 46.760655959447234s)  
 Loss: 0.0926(0.1166)  
 EVAL: [40/62] Elapsed 0m 46.85940146446228s (remain 0m 24.001156847651416s)  
 Loss: 0.1071(0.1150)  
 EVAL: [60/62] Elapsed 1m 12.76200819015503s (remain 0m 1.1928198063959883s)  
 Loss: 0.1008(0.1160)

EVAL: [61/62] Elapsed 1m 12.814218521118164s (remain 0m 0.0s) Loss:  
 0.0581(0.1159)  
 Epoch 2 - avg\_train\_loss: 0.1027 avg\_val\_loss: 0.1159 time: 706s  
 Epoch 2 - avg\_train\_loss: 0.1027 avg\_val\_loss: 0.1159 time: 706s  
 Epoch 2 - avg\_train\_loss: 0.1027 avg\_val\_loss: 0.1159 time: 706s  
 Epoch 2 - Score: 0.4825 Scores: [0.5426077268785788, 0.48342925133878145,  
 0.43830955705585667, 0.4830111980068668, 0.4815850441579428,  
 0.46596111457071987]  
 Epoch 2 - Score: 0.4825 Scores: [0.5426077268785788, 0.48342925133878145,  
 0.43830955705585667, 0.4830111980068668, 0.4815850441579428,  
 0.46596111457071987]  
 Epoch 2 - Score: 0.4825 Scores: [0.5426077268785788, 0.48342925133878145,  
 0.43830955705585667, 0.4830111980068668, 0.4815850441579428,  
 0.46596111457071987]  
 Epoch 2 - Save Best Score: 0.4825 Model  
 Epoch 2 - Save Best Score: 0.4825 Model  
 Epoch 2 - Save Best Score: 0.4825 Model  
 Epoch: [3][0/366] Elapsed 0m 1.3623402118682861s (remain 8m 17.25417733192444s)  
 Loss: 0.0596(0.0596) Grad: 211324.8594 LR: 0.00001001  
 Epoch: [3][20/366] Elapsed 0m 32.452213764190674s (remain 8m 53.14351184027532s)  
 Loss: 0.0787(0.0906) Grad: 202719.1719 LR: 0.00000958  
 Epoch: [3][40/366] Elapsed 1m 7.358182907104492s (remain 8m 53.936815727047815s)  
 Loss: 0.0715(0.0934) Grad: 201269.8281 LR: 0.00000916  
 Epoch: [3][60/366] Elapsed 1m 45.35917091369629s (remain 8m 46.795854568481445s)  
 Loss: 0.0707(0.0917) Grad: 189779.2656 LR: 0.00000873  
 Epoch: [3][80/366] Elapsed 2m 22.3607394695282s (remain 8m 20.89889813352511s)  
 Loss: 0.0927(0.0922) Grad: 116567.1719 LR: 0.00000831  
 Epoch: [3][100/366] Elapsed 2m 52.71647834777832s (remain 7m 33.16699764516102s)  
 Loss: 0.1050(0.0915) Grad: 92266.1094 LR: 0.00000789  
 Epoch: [3][120/366] Elapsed 3m 24.871464252471924s (remain 6m  
 54.82238629632752s) Loss: 0.0991(0.0907) Grad: 93794.5625 LR: 0.00000747  
 Epoch: [3][140/366] Elapsed 3m 57.210612535476685s (remain 6m  
 18.527573194909678s) Loss: 0.0663(0.0889) Grad: 62603.7656 LR: 0.00000706  
 Epoch: [3][160/366] Elapsed 4m 32.10998344421387s (remain 5m  
 46.475444758160506s) Loss: 0.1379(0.0912) Grad: 88968.0625 LR: 0.00000665  
 Epoch: [3][180/366] Elapsed 5m 1.5259640216827393s (remain 5m  
 8.189521237631538s) Loss: 0.1283(0.0913) Grad: 126034.0000 LR: 0.00000625  
 Epoch: [3][200/366] Elapsed 5m 35.50549054145813s (remain 4m 35.41495492209242s)  
 Loss: 0.0862(0.0921) Grad: 79877.7578 LR: 0.00000586  
 Epoch: [3][220/366] Elapsed 6m 14.472579002380371s (remain 4m  
 5.694678530973533s) Loss: 0.0977(0.0927) Grad: 220555.4844 LR: 0.00000547  
 Epoch: [3][240/366] Elapsed 6m 50.70072889328003s (remain 3m 33.01905025585063s)  
 Loss: 0.0866(0.0934) Grad: 114538.7812 LR: 0.00000509  
 Epoch: [3][260/366] Elapsed 7m 19.680066347122192s (remain 2m  
 56.882785312060605s) Loss: 0.0704(0.0932) Grad: 113999.6641 LR: 0.00000472  
 Epoch: [3][280/366] Elapsed 8m 3.4590861797332764s (remain 2m  
 26.24207233194784s) Loss: 0.0651(0.0928) Grad: 71958.8125 LR: 0.00000437  
 Epoch: [3][300/366] Elapsed 8m 39.5053014755249s (remain 1m 52.18553021896719s)

Loss: 0.1073(0.0932) Grad: 100844.1875 LR: 0.00000402  
 Epoch: [3][320/366] Elapsed 9m 10.406760454177856s (remain 1m 17.159826231894158s) Loss: 0.0619(0.0933) Grad: 100751.6016 LR: 0.00000368  
 Epoch: [3][340/366] Elapsed 9m 47.869332790374756s (remain 0m 43.098924691376396s) Loss: 0.0931(0.0937) Grad: 101327.7500 LR: 0.00000335  
 Epoch: [3][360/366] Elapsed 10m 19.243282318115234s (remain 0m 8.576776763408816s) Loss: 0.1099(0.0936) Grad: 121593.6641 LR: 0.00000304  
 Epoch: [3][365/366] Elapsed 10m 27.33644986152649s (remain 0m 0.0s) Loss: 0.1000(0.0937) Grad: 74047.7578 LR: 0.00000296  
 EVAL: [0/62] Elapsed 0m 0.8731734752655029s (remain 0m 53.26358199119568s) Loss: 0.0987(0.0987)  
 EVAL: [20/62] Elapsed 0m 23.945639848709106s (remain 0m 46.75101113319397s) Loss: 0.1087(0.1102)  
 EVAL: [40/62] Elapsed 0m 46.84661245346069s (remain 0m 23.994606378601816s) Loss: 0.0964(0.1061)  
 EVAL: [60/62] Elapsed 1m 12.74852180480957s (remain 0m 1.1925987181116255s) Loss: 0.0895(0.1045)  
 EVAL: [61/62] Elapsed 1m 12.800435543060303s (remain 0m 0.0s) Loss: 0.0530(0.1045)  
 Epoch 3 - avg\_train\_loss: 0.0937 avg\_val\_loss: 0.1045 time: 700s  
 Epoch 3 - avg\_train\_loss: 0.0937 avg\_val\_loss: 0.1045 time: 700s  
 Epoch 3 - avg\_train\_loss: 0.0937 avg\_val\_loss: 0.1045 time: 700s  
 Epoch 3 - Score: 0.4575 Scores: [0.49528149426944024, 0.44241485276759346, 0.41686757047750045, 0.4496009522133593, 0.4866076918808725, 0.4544972587706356]  
 Epoch 3 - Score: 0.4575 Scores: [0.49528149426944024, 0.44241485276759346, 0.41686757047750045, 0.4496009522133593, 0.4866076918808725, 0.4544972587706356]  
 Epoch 3 - Score: 0.4575 Scores: [0.49528149426944024, 0.44241485276759346, 0.41686757047750045, 0.4496009522133593, 0.4866076918808725, 0.4544972587706356]  
 Epoch 3 - Save Best Score: 0.4575 Model  
 Epoch 3 - Save Best Score: 0.4575 Model  
 Epoch 3 - Save Best Score: 0.4575 Model  
 Epoch: [4][0/366] Elapsed 0m 2.201040744781494s (remain 13m 23.37987184524536s) Loss: 0.0897(0.0897) Grad: 132290.7031 LR: 0.00000295  
 Epoch: [4][20/366] Elapsed 0m 36.39856576919556s (remain 9m 57.97643763678411s) Loss: 0.1013(0.0903) Grad: 121613.5547 LR: 0.00000265  
 Epoch: [4][40/366] Elapsed 1m 15.977033615112305s (remain 10m 2.256973778329211s) Loss: 0.0605(0.0909) Grad: 85666.9688 LR: 0.00000237  
 Epoch: [4][60/366] Elapsed 1m 49.678631067276s (remain 9m 8.393155336380005s) Loss: 0.0558(0.0914) Grad: 91270.6328 LR: 0.00000210  
 Epoch: [4][80/366] Elapsed 2m 23.838181972503662s (remain 8m 26.09730694029065s) Loss: 0.0860(0.0905) Grad: 65735.7266 LR: 0.00000184  
 Epoch: [4][100/366] Elapsed 2m 53.070109844207764s (remain 7m 34.09484266054517s) Loss: 0.0925(0.0909) Grad: 106448.0312 LR: 0.00000160  
 Epoch: [4][120/366] Elapsed 3m 31.78116512298584s (remain 7m 8.813102934971425s) Loss: 0.0631(0.0912) Grad: 88830.5703 LR: 0.00000138  
 Epoch: [4][140/366] Elapsed 4m 2.053762435913086s (remain 6m 26.256003887095403s) Loss: 0.1526(0.0911) Grad: 178534.0312 LR: 0.00000117  
 Epoch: [4][160/366] Elapsed 4m 37.25459694862366s (remain 5m

53.026039592968004s) Loss: 0.0861(0.0909) Grad: 62802.6406 LR: 0.00000098  
Epoch: [4][180/366] Elapsed 5m 13.787365674972534s (remain 5m  
20.721893093204017s) Loss: 0.0951(0.0908) Grad: 63092.0625 LR: 0.00000080  
Epoch: [4][200/366] Elapsed 5m 47.259899854660034s (remain 4m  
45.06409689561633s) Loss: 0.1292(0.0905) Grad: 100232.6719 LR: 0.00000064  
Epoch: [4][220/366] Elapsed 6m 26.56472611427307s (remain 4m  
13.628440210722147s) Loss: 0.0484(0.0900) Grad: 59673.8516 LR: 0.00000050  
Epoch: [4][240/366] Elapsed 7m 0.9086906909942627s (remain 3m  
38.31363625051574s) Loss: 0.0961(0.0899) Grad: 63258.8164 LR: 0.00000037  
Epoch: [4][260/366] Elapsed 7m 34.02416801452637s (remain 3m 2.653400925384176s)  
Loss: 0.0587(0.0897) Grad: 51040.2070 LR: 0.00000027  
Epoch: [4][280/366] Elapsed 8m 9.658151626586914s (remain 2m  
28.117234477793204s) Loss: 0.0842(0.0892) Grad: 137029.5469 LR: 0.00000018  
Epoch: [4][300/366] Elapsed 8m 45.00165772438049s (remain 1m  
53.372451003603715s) Loss: 0.0816(0.0889) Grad: 81505.0781 LR: 0.00000011  
Epoch: [4][320/366] Elapsed 9m 13.40463638305664s (remain 1m 17.58008921257806s)  
Loss: 0.0820(0.0891) Grad: 53225.7656 LR: 0.00000005  
Epoch: [4][340/366] Elapsed 9m 46.43552041053772s (remain 0m  
42.993806481710976s) Loss: 0.0875(0.0884) Grad: 88326.3516 LR: 0.00000002  
Epoch: [4][360/366] Elapsed 10m 18.3008930683136s (remain 0m 8.563724280724614s)  
Loss: 0.0943(0.0884) Grad: 99200.6641 LR: 0.00000000  
Epoch: [4][365/366] Elapsed 10m 26.72061562538147s (remain 0m 0.0s) Loss:  
0.1371(0.0884) Grad: 152874.5625 LR: 0.00000000  
EVAL: [0/62] Elapsed 0m 0.8934314250946045s (remain 0m 54.499316930770874s)  
Loss: 0.1051(0.1051)  
EVAL: [20/62] Elapsed 0m 23.966991424560547s (remain 0m 46.79269754318965s)  
Loss: 0.1001(0.1086)  
EVAL: [40/62] Elapsed 0m 46.86725401878357s (remain 0m 24.005178887669643s)  
Loss: 0.1004(0.1051)  
EVAL: [60/62] Elapsed 1m 12.76876163482666s (remain 0m 1.192930518603717s) Loss:  
0.0931(0.1045)  
EVAL: [61/62] Elapsed 1m 12.821126937866211s (remain 0m 0.0s) Loss:  
0.0477(0.1045)  
Epoch 4 - avg\_train\_loss: 0.0884 avg\_val\_loss: 0.1045 time: 700s  
Epoch 4 - avg\_train\_loss: 0.0884 avg\_val\_loss: 0.1045 time: 700s  
Epoch 4 - avg\_train\_loss: 0.0884 avg\_val\_loss: 0.1045 time: 700s  
Epoch 4 - Score: 0.4576 Scores: [0.49642531090253766, 0.444601057736994,  
0.41759681937050885, 0.4516415532491227, 0.48101894390053945,  
0.45441870643587917]  
Epoch 4 - Score: 0.4576 Scores: [0.49642531090253766, 0.444601057736994,  
0.41759681937050885, 0.4516415532491227, 0.48101894390053945,  
0.45441870643587917]  
Epoch 4 - Score: 0.4576 Scores: [0.49642531090253766, 0.444601057736994,  
0.41759681937050885, 0.4516415532491227, 0.48101894390053945,  
0.45441870643587917]  
----- Fold: 1 Result -----  
----- Fold: 1 Result -----  
----- Fold: 1 Result -----



Score: 0.4575 Scores: [0.49528149426944024, 0.44241485276759346,  
0.41686757047750045, 0.4496009522133593, 0.4866076918808725, 0.4544972587706356]  
Score: 0.4575 Scores: [0.49528149426944024, 0.44241485276759346,  
0.41686757047750045, 0.4496009522133593, 0.4866076918808725, 0.4544972587706356]  
Score: 0.4575 Scores: [0.49528149426944024, 0.44241485276759346,  
0.41686757047750045, 0.4496009522133593, 0.4866076918808725, 0.4544972587706356]

----- Fold: 2 Training -----  
----- Fold: 2 Training -----  
----- Fold: 2 Training -----

```
DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",
  "hidden_dropout": 0.0,
  "hidden_dropout_prob": 0.0,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-07,
  "max_position_embeddings": 512,
  "max_relative_positions": -1,
  "model_type": "deberta-v2",
  "norm_rel_ebd": "layer_norm",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_hidden_states": true,
  "pad_token_id": 0,
  "pooler_dropout": 0,
  "pooler_hidden_act": "gelu",
  "pooler_hidden_size": 768,
  "pos_att_type": [
    "p2c",
    "c2p"
  ],
  "position_biased_input": false,
  "position_buckets": 256,
  "relative_attention": true,
  "share_att_key": true,
  "transformers_version": "4.42.3",
  "type_vocab_size": 0,
  "vocab_size": 128100
}
```

```
DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
```

```

"hidden_act": "gelu",
"hidden_dropout": 0.0,
"hidden_dropout_prob": 0.0,
"hidden_size": 768,
"initializer_range": 0.02,
"intermediate_size": 3072,
"layer_norm_eps": 1e-07,
"max_position_embeddings": 512,
"max_relative_positions": -1,
"model_type": "deberta-v2",
"norm_rel_ebd": "layer_norm",
"num_attention_heads": 12,
"num_hidden_layers": 12,
"output_hidden_states": true,
"pad_token_id": 0,
"pooler_dropout": 0,
"pooler_hidden_act": "gelu",
"pooler_hidden_size": 768,
"pos_att_type": [
    "p2c",
    "c2p"
],
"position_biased_input": false,
"position_buckets": 256,
"relative_attention": true,
"share_att_key": true,
"transformers_version": "4.42.3",
"type_vocab_size": 0,
"vocab_size": 128100
}

DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",
  "hidden_dropout": 0.0,
  "hidden_dropout_prob": 0.0,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-07,
  "max_position_embeddings": 512,
  "max_relative_positions": -1,
  "model_type": "deberta-v2",
  "norm_rel_ebd": "layer_norm",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,

```

```

"output_hidden_states": true,
"pad_token_id": 0,
"pooler_dropout": 0,
"pooler_hidden_act": "gelu",
"pooler_hidden_size": 768,
"pos_att_type": [
    "p2c",
    "c2p"
],
"position_biased_input": false,
"position_buckets": 256,
"relative_attention": true,
"share_att_key": true,
"transformers_version": "4.42.3",
"type_vocab_size": 0,
"vocab_size": 128100
}

```

```

Epoch: [1][0/366] Elapsed 0m 2.2315409183502197s (remain 13m 34.5124351978302s)
Loss: 3.0163(3.0163) Grad: inf LR: 0.00002000
Epoch: [1][20/366] Elapsed 0m 35.163010120391846s (remain 9m 37.67802340643743s)
Loss: 0.4737(1.8034) Grad: 144692.6562 LR: 0.00001999
Epoch: [1][40/366] Elapsed 1m 9.468846321105957s (remain 9m 10.667684252669119s)
Loss: 0.2834(1.0197) Grad: 261468.3281 LR: 0.00001996
Epoch: [1][60/366] Elapsed 1m 45.35833287239075s (remain 8m 46.791664361953735s)
Loss: 0.1675(0.7386) Grad: 94055.9766 LR: 0.00001991
Epoch: [1][80/366] Elapsed 2m 17.870519876480103s (remain 8m 5.099977343170735s)
Loss: 0.0748(0.5926) Grad: 61508.1445 LR: 0.00001985
Epoch: [1][100/366] Elapsed 2m 50.151323080062866s (remain 7m
26.436639764521487s) Loss: 0.1104(0.5034) Grad: 42208.4648 LR: 0.00001977
Epoch: [1][120/366] Elapsed 3m 27.50582003593445s (remain 7m
0.15641246945403964s) Loss: 0.1111(0.4424) Grad: 73225.1641 LR: 0.00001967
Epoch: [1][140/366] Elapsed 4m 6.212651968002319s (remain 6m 32.89252973617397s)
Loss: 0.0649(0.3984) Grad: 46792.6953 LR: 0.00001955
Epoch: [1][160/366] Elapsed 4m 46.15047478675842s (remain 6m 4.353089014195575s)
Loss: 0.0772(0.3663) Grad: 39096.7305 LR: 0.00001941
Epoch: [1][180/366] Elapsed 5m 23.228686809539795s (remain 5m
30.371862208645666s) Loss: 0.0740(0.3408) Grad: 68590.5781 LR: 0.00001926
Epoch: [1][200/366] Elapsed 5m 57.47189784049988s (remain 4m 53.44708031682819s)
Loss: 0.0823(0.3199) Grad: 43522.7617 LR: 0.00001909
Epoch: [1][220/366] Elapsed 6m 36.42440605163574s (remain 4m 20.09746098410494s)
Loss: 0.1094(0.3031) Grad: 41520.2227 LR: 0.00001890
Epoch: [1][240/366] Elapsed 7m 7.754178524017334s (remain 3m
41.864200479262195s) Loss: 0.0963(0.2870) Grad: 54427.1758 LR: 0.00001870
Epoch: [1][260/366] Elapsed 7m 42.07360911369324s (remain 3m 5.891681827347838s)
Loss: 0.0838(0.2747) Grad: 40539.4219 LR: 0.00001848
Epoch: [1][280/366] Elapsed 8m 18.21062922477722s (remain 2m
30.704282861587444s) Loss: 0.1399(0.2665) Grad: 50529.2148 LR: 0.00001824

```

Epoch: [1][300/366] Elapsed 8m 50.93566632270813s (remain 1m 54.65388143181406s)  
 Loss: 0.1230(0.2565) Grad: 60350.1914 LR: 0.00001799  
 Epoch: [1][320/366] Elapsed 9m 25.31002640724182s (remain 1m 19.249069122510605s) Loss: 0.0866(0.2473) Grad: 41140.8750 LR: 0.00001773  
 Epoch: [1][340/366] Elapsed 10m 3.490895986557007s (remain 0m 44.24420058552471s) Loss: 0.1134(0.2408) Grad: 89462.8906 LR: 0.00001745  
 Epoch: [1][360/366] Elapsed 10m 41.033228635787964s (remain 0m 8.878576573902933s) Loss: 0.1049(0.2339) Grad: 70761.6641 LR: 0.00001715  
 Epoch: [1][365/366] Elapsed 10m 50.00003385543823s (remain 0m 0.0s) Loss: 0.1130(0.2324) Grad: 45747.5273 LR: 0.00001708  
 EVAL: [0/62] Elapsed 0m 2.0964479446411133s (remain 2m 7.88332462310791s) Loss: 0.1356(0.1356)  
 EVAL: [20/62] Elapsed 0m 23.81535530090332s (remain 0m 46.49664606366839s) Loss: 0.0871(0.1211)  
 EVAL: [40/62] Elapsed 0m 43.81618785858154s (remain 0m 22.442437683663726s) Loss: 0.0901(0.1191)  
 EVAL: [60/62] Elapsed 1m 5.917513608932495s (remain 0m 1.0806149771956086s) Loss: 0.0980(0.1193)  
 EVAL: [61/62] Elapsed 1m 6.00235652923584s (remain 0m 0.0s) Loss: 0.1484(0.1194)  
 Epoch 1 - avg\_train\_loss: 0.2324 avg\_val\_loss: 0.1194 time: 716s  
 Epoch 1 - avg\_train\_loss: 0.2324 avg\_val\_loss: 0.1194 time: 716s  
 Epoch 1 - avg\_train\_loss: 0.2324 avg\_val\_loss: 0.1194 time: 716s  
 Epoch 1 - Score: 0.4901 Scores: [0.5272298538395546, 0.49310997713540705, 0.44135011515728445, 0.49237477552348874, 0.5119441931336799, 0.4745026189151668]  
 Epoch 1 - Score: 0.4901 Scores: [0.5272298538395546, 0.49310997713540705, 0.44135011515728445, 0.49237477552348874, 0.5119441931336799, 0.4745026189151668]  
 Epoch 1 - Score: 0.4901 Scores: [0.5272298538395546, 0.49310997713540705, 0.44135011515728445, 0.49237477552348874, 0.5119441931336799, 0.4745026189151668]  
 Epoch 1 - Save Best Score: 0.4901 Model  
 Epoch 1 - Save Best Score: 0.4901 Model  
 Epoch 1 - Save Best Score: 0.4901 Model  
 Epoch: [2][0/366] Elapsed 0m 1.5856316089630127s (remain 9m 38.755537271499634s) Loss: 0.1186(0.1186) Grad: 154034.4375 LR: 0.00001706  
 Epoch: [2][20/366] Elapsed 0m 34.65964460372925s (remain 9m 29.4084470612662s) Loss: 0.1228(0.1261) Grad: 96660.0469 LR: 0.00001675  
 Epoch: [2][40/366] Elapsed 1m 11.516374826431274s (remain 9m 26.898093136345437s) Loss: 0.1472(0.1228) Grad: 115961.7422 LR: 0.00001643  
 Epoch: [2][60/366] Elapsed 1m 42.97416043281555s (remain 8m 34.87080216407776s) Loss: 0.0899(0.1157) Grad: 46976.5820 LR: 0.00001610  
 Epoch: [2][80/366] Elapsed 2m 22.90174674987793s (remain 8m 22.802442268089067s) Loss: 0.0884(0.1129) Grad: 35086.3672 LR: 0.00001575  
 Epoch: [2][100/366] Elapsed 2m 52.964778661727905s (remain 7m 33.81847866690987s) Loss: 0.1635(0.1130) Grad: 68619.7031 LR: 0.00001540  
 Epoch: [2][120/366] Elapsed 3m 28.780233144760132s (remain 7m 2.7368357063324993s) Loss: 0.1112(0.1118) Grad: 86347.2109 LR: 0.00001503

Epoch: [2][140/366] Elapsed 4m 0.8580019474029541s (remain 6m 24.347875447983483s) Loss: 0.1495(0.1108) Grad: 48499.4414 LR: 0.00001466

Epoch: [2][160/366] Elapsed 4m 33.19124507904053s (remain 5m 47.85220646710127s) Loss: 0.0695(0.1111) Grad: 27199.3359 LR: 0.00001427

Epoch: [2][180/366] Elapsed 5m 3.4487154483795166s (remain 5m 10.154764408564688s) Loss: 0.0953(0.1109) Grad: 40730.8281 LR: 0.00001388

Epoch: [2][200/366] Elapsed 5m 43.80561947822571s (remain 4m 42.2284936015285s) Loss: 0.1469(0.1109) Grad: 65172.1016 LR: 0.00001348

Epoch: [2][220/366] Elapsed 6m 23.13970446586609s (remain 4m 11.381254061314849s) Loss: 0.0675(0.1108) Grad: 69009.7344 LR: 0.00001308

Epoch: [2][240/366] Elapsed 6m 56.485026836395264s (remain 3m 36.01920479066143s) Loss: 0.1082(0.1105) Grad: 38414.5703 LR: 0.00001267

Epoch: [2][260/366] Elapsed 7m 35.20084190368652s (remain 3m 3.126775478494551s) Loss: 0.0769(0.1106) Grad: 60778.8789 LR: 0.00001225

Epoch: [2][280/366] Elapsed 8m 13.956336498260498s (remain 2m 29.41739716139557s) Loss: 0.0922(0.1106) Grad: 62956.3320 LR: 0.00001183

Epoch: [2][300/366] Elapsed 8m 47.45966076850891s (remain 1m 53.903249003166366s) Loss: 0.0865(0.1115) Grad: 47713.2266 LR: 0.00001141

Epoch: [2][320/366] Elapsed 9m 19.49643898010254s (remain 1m 18.434080230855557s) Loss: 0.1578(0.1110) Grad: 77873.2578 LR: 0.00001098

Epoch: [2][340/366] Elapsed 9m 53.669129610061646s (remain 0m 43.52412973680805s) Loss: 0.1281(0.1110) Grad: 38261.8281 LR: 0.00001056

Epoch: [2][360/366] Elapsed 10m 36.66359877586365s (remain 0m 8.818055384707236s) Loss: 0.1067(0.1102) Grad: 31067.5762 LR: 0.00001013

Epoch: [2][365/366] Elapsed 10m 45.21775579452515s (remain 0m 0.0s) Loss: 0.0716(0.1101) Grad: 47429.3438 LR: 0.00001002

EVAL: [0/62] Elapsed 0m 1.9001519680023193s (remain 1m 55.90927004814148s) Loss: 0.1448(0.1448)

EVAL: [20/62] Elapsed 0m 23.206724166870117s (remain 0m 45.30836623055595s) Loss: 0.0916(0.1115)

EVAL: [40/62] Elapsed 0m 42.98834824562073s (remain 0m 22.018422272147205s) Loss: 0.0943(0.1098)

EVAL: [60/62] Elapsed 1m 4.899736642837524s (remain 0m 1.0639301088989725s) Loss: 0.1012(0.1106)

EVAL: [61/62] Elapsed 1m 4.984724283218384s (remain 0m 0.0s) Loss: 0.1270(0.1106)

Epoch 2 - avg\_train\_loss: 0.1101 avg\_val\_loss: 0.1106 time: 710s

Epoch 2 - avg\_train\_loss: 0.1101 avg\_val\_loss: 0.1106 time: 710s

Epoch 2 - avg\_train\_loss: 0.1101 avg\_val\_loss: 0.1106 time: 710s

Epoch 2 - Score: 0.4717 Scores: [0.49790621415952147, 0.46080148694660716, 0.4275591443671682, 0.4770468360825199, 0.4955204861468956, 0.4715107833995943]

Epoch 2 - Score: 0.4717 Scores: [0.49790621415952147, 0.46080148694660716, 0.4275591443671682, 0.4770468360825199, 0.4955204861468956, 0.4715107833995943]

Epoch 2 - Score: 0.4717 Scores: [0.49790621415952147, 0.46080148694660716, 0.4275591443671682, 0.4770468360825199, 0.4955204861468956, 0.4715107833995943]

Epoch 2 - Save Best Score: 0.4717 Model

Epoch 2 - Save Best Score: 0.4717 Model

Epoch 2 - Save Best Score: 0.4717 Model

Epoch: [3][0/366] Elapsed 0m 1.1070365905761719s (remain 6m 44.068355560302734s)  
 Loss: 0.0917(0.0917) Grad: 112808.0547 LR: 0.00001000  
 Epoch: [3][20/366] Elapsed 0m 38.25712013244629s (remain 10m 28.509830747331876s) Loss: 0.1000(0.0994) Grad: 182652.6250 LR: 0.00000957  
 Epoch: [3][40/366] Elapsed 1m 10.896766901016235s (remain 9m 21.98656689829943s)  
 Loss: 0.0947(0.0994) Grad: 85043.0391 LR: 0.00000914  
 Epoch: [3][60/366] Elapsed 1m 48.055217266082764s (remain 9m 0.27608633041381836s) Loss: 0.0755(0.1040) Grad: 73881.5391 LR: 0.00000872  
 Epoch: [3][80/366] Elapsed 2m 19.63500189781189s (remain 8m 11.30834001081962s)  
 Loss: 0.1092(0.1029) Grad: 42013.8711 LR: 0.00000829  
 Epoch: [3][100/366] Elapsed 2m 55.92603659629822s (remain 7m 41.58811582197063s)  
 Loss: 0.1084(0.1004) Grad: 147053.1719 LR: 0.00000787  
 Epoch: [3][120/366] Elapsed 3m 36.21222758293152s (remain 7m 17.7850889075886s)  
 Loss: 0.0704(0.0989) Grad: 82413.2812 LR: 0.00000746  
 Epoch: [3][140/366] Elapsed 4m 9.331796169281006s (remain 6m 37.86988750417186s)  
 Loss: 0.0837(0.0999) Grad: 45290.8203 LR: 0.00000704  
 Epoch: [3][160/366] Elapsed 4m 46.44657921791077s (remain 6m 4.730116395476443s)  
 Loss: 0.0858(0.0996) Grad: 137723.5781 LR: 0.00000664  
 Epoch: [3][180/366] Elapsed 5m 24.0592041015625s (remain 5m 31.2207334739727s)  
 Loss: 0.1043(0.0986) Grad: 145773.6562 LR: 0.00000624  
 Epoch: [3][200/366] Elapsed 6m 1.7810859680175781s (remain 4m 56.98447355583528s) Loss: 0.0852(0.0981) Grad: 64154.0703 LR: 0.00000584  
 Epoch: [3][220/366] Elapsed 6m 35.3042311668396s (remain 4m 19.36250461172733s)  
 Loss: 0.0525(0.0973) Grad: 81528.5391 LR: 0.00000546  
 Epoch: [3][240/366] Elapsed 7m 9.450932264328003s (remain 3m 42.744259473199236s) Loss: 0.0619(0.0959) Grad: 155126.8438 LR: 0.00000508  
 Epoch: [3][260/366] Elapsed 7m 39.44951367378235s (remain 3m 4.836011248073305s)  
 Loss: 0.1191(0.0959) Grad: 102255.0391 LR: 0.00000471  
 Epoch: [3][280/366] Elapsed 8m 17.277354955673218s (remain 2m 30.42197569833536s) Loss: 0.1086(0.0956) Grad: 102062.3047 LR: 0.00000435  
 Epoch: [3][300/366] Elapsed 8m 50.47245121002197s (remain 1m 54.55385159020409s)  
 Loss: 0.1053(0.0962) Grad: 156358.7812 LR: 0.00000400  
 Epoch: [3][320/366] Elapsed 9m 32.38995718955994s (remain 1m 20.241582783583226s) Loss: 0.0515(0.0958) Grad: 80230.5000 LR: 0.00000367  
 Epoch: [3][340/366] Elapsed 10m 9.501128196716309s (remain 0m 44.684833445507024s) Loss: 0.1101(0.0953) Grad: 60081.0977 LR: 0.00000334  
 Epoch: [3][360/366] Elapsed 10m 42.31630325317383s (remain 0m 8.896347690487232s) Loss: 0.0963(0.0954) Grad: 89378.2266 LR: 0.00000303  
 Epoch: [3][365/366] Elapsed 10m 52.451236963272095s (remain 0m 0.0s) Loss: 0.1089(0.0953) Grad: 100604.3359 LR: 0.00000295  
 EVAL: [0/62] Elapsed 0m 1.9040067195892334s (remain 1m 56.14440989494324s) Loss: 0.1340(0.1340)  
 EVAL: [20/62] Elapsed 0m 23.21494722366333s (remain 0m 45.32442077000937s) Loss: 0.0893(0.1090)  
 EVAL: [40/62] Elapsed 0m 42.99456739425659s (remain 0m 22.021607689741188s) Loss: 0.0906(0.1072)  
 EVAL: [60/62] Elapsed 1m 4.907400131225586s (remain 0m 1.0640557398561583s) Loss: 0.1010(0.1080)

EVAL: [61/62] Elapsed 1m 4.992188453674316s (remain 0m 0.0s) Loss:  
 0.1346(0.1080)  
 Epoch 3 - avg\_train\_loss: 0.0953 avg\_val\_loss: 0.1080 time: 718s  
 Epoch 3 - avg\_train\_loss: 0.0953 avg\_val\_loss: 0.1080 time: 718s  
 Epoch 3 - avg\_train\_loss: 0.0953 avg\_val\_loss: 0.1080 time: 718s  
 Epoch 3 - Score: 0.4659 Scores: [0.49470072662870057, 0.45568513873210703,  
 0.42256914320312633, 0.47108572264160514, 0.48980836646349746,  
 0.46166749538934576]  
 Epoch 3 - Score: 0.4659 Scores: [0.49470072662870057, 0.45568513873210703,  
 0.42256914320312633, 0.47108572264160514, 0.48980836646349746,  
 0.46166749538934576]  
 Epoch 3 - Score: 0.4659 Scores: [0.49470072662870057, 0.45568513873210703,  
 0.42256914320312633, 0.47108572264160514, 0.48980836646349746,  
 0.46166749538934576]  
 Epoch 3 - Save Best Score: 0.4659 Model  
 Epoch 3 - Save Best Score: 0.4659 Model  
 Epoch 3 - Save Best Score: 0.4659 Model  
 Epoch: [4][0/366] Elapsed 0m 1.2869477272033691s (remain 7m 49.735920429229736s)  
 Loss: 0.0722(0.0722) Grad: 215231.1719 LR: 0.00000294  
 Epoch: [4][20/366] Elapsed 0m 38.21220874786377s (remain 10m  
 27.772000857761896s) Loss: 0.0642(0.0929) Grad: 93828.8438 LR: 0.00000264  
 Epoch: [4][40/366] Elapsed 1m 13.302214622497559s (remain 9m 41.05414030028555s)  
 Loss: 0.0779(0.0891) Grad: 157816.5469 LR: 0.00000236  
 Epoch: [4][60/366] Elapsed 1m 45.791391372680664s (remain 8m 48.95695686340332s)  
 Loss: 0.0883(0.0877) Grad: 75169.9297 LR: 0.00000209  
 Epoch: [4][80/366] Elapsed 2m 24.945001125335693s (remain 8m 29.99167062618119s)  
 Loss: 0.1177(0.0880) Grad: 145913.5000 LR: 0.00000183  
 Epoch: [4][100/366] Elapsed 3m 2.483726739883423s (remain 7m 58.79393649573376s)  
 Loss: 0.0572(0.0876) Grad: 57378.5234 LR: 0.00000159  
 Epoch: [4][120/366] Elapsed 3m 35.20608711242676s (remain 7m  
 15.747862335079049s) Loss: 0.0964(0.0890) Grad: 138226.9375 LR: 0.00000137  
 Epoch: [4][140/366] Elapsed 4m 14.837924003601074s (remain 6m  
 46.65626170787414s) Loss: 0.0785(0.0905) Grad: 186098.8594 LR: 0.00000116  
 Epoch: [4][160/366] Elapsed 4m 52.2081503868103s (remain 6m 12.066278442832981s)  
 Loss: 0.0655(0.0894) Grad: 75707.4531 LR: 0.00000097  
 Epoch: [4][180/366] Elapsed 5m 24.04688310623169s (remain 5m  
 31.208140191452344s) Loss: 0.0690(0.0890) Grad: 75527.4062 LR: 0.00000079  
 Epoch: [4][200/366] Elapsed 6m 1.3373327255249023s (remain 4m  
 56.620198506027805s) Loss: 0.0719(0.0895) Grad: 54567.1172 LR: 0.00000063  
 Epoch: [4][220/366] Elapsed 6m 40.339574098587036s (remain 4m  
 22.666236399525474s) Loss: 0.1050(0.0900) Grad: 84575.0938 LR: 0.00000049  
 Epoch: [4][240/366] Elapsed 7m 16.064406156539917s (remain 3m  
 46.17448452102701s) Loss: 0.0881(0.0893) Grad: 40953.7500 LR: 0.00000037  
 Epoch: [4][260/366] Elapsed 7m 49.721900939941406s (remain 3m  
 8.968580837907439s) Loss: 0.1036(0.0891) Grad: 92997.4141 LR: 0.00000026  
 Epoch: [4][280/366] Elapsed 8m 22.125609159469604s (remain 2m  
 31.88852946104953s) Loss: 0.0907(0.0889) Grad: 140360.5000 LR: 0.00000017  
 Epoch: [4][300/366] Elapsed 8m 52.322869300842285s (remain 1m

```

54.953443536726695s) Loss: 0.0809(0.0886) Grad: 138470.5312 LR: 0.00000010
Epoch: [4][320/366] Elapsed 9m 27.131654977798462s (remain 1m
19.504437613710138s) Loss: 0.1013(0.0887) Grad: 68231.4609 LR: 0.00000005
Epoch: [4][340/366] Elapsed 10m 2.56792950630188s (remain 0m 44.17653442128312s)
Loss: 0.0757(0.0886) Grad: 73569.0625 LR: 0.00000002
Epoch: [4][360/366] Elapsed 10m 38.40600919723511s (remain 0m
8.842188493036474s) Loss: 0.1143(0.0885) Grad: 93339.6484 LR: 0.00000000
Epoch: [4][365/366] Elapsed 10m 45.43635606765747s (remain 0m 0.0s) Loss:
0.0860(0.0887) Grad: 74801.8203 LR: 0.00000000
EVAL: [0/62] Elapsed 0m 1.8830888271331787s (remain 1m 54.8684184551239s) Loss:
0.1321(0.1321)
EVAL: [20/62] Elapsed 0m 23.195310354232788s (remain 0m 45.286082120168786s)
Loss: 0.0937(0.1088)
EVAL: [40/62] Elapsed 0m 42.973031520843506s (remain 0m 22.010577120432046s)
Loss: 0.0917(0.1070)
EVAL: [60/62] Elapsed 1m 4.881884813308716s (remain 0m 1.0636374559558845s)
Loss: 0.1007(0.1079)
EVAL: [61/62] Elapsed 1m 4.966566324234009s (remain 0m 0.0s) Loss:
0.1487(0.1079)
Epoch 4 - avg_train_loss: 0.0887 avg_val_loss: 0.1079 time: 711s
Epoch 4 - avg_train_loss: 0.0887 avg_val_loss: 0.1079 time: 711s
Epoch 4 - avg_train_loss: 0.0887 avg_val_loss: 0.1079 time: 711s
Epoch 4 - Score: 0.4657 Scores: [0.49372475227573576, 0.45629093313936075,
0.4215764865653166, 0.4703336143410078, 0.4900214306483105, 0.46236955586870565]
Epoch 4 - Score: 0.4657 Scores: [0.49372475227573576, 0.45629093313936075,
0.4215764865653166, 0.4703336143410078, 0.4900214306483105, 0.46236955586870565]
Epoch 4 - Score: 0.4657 Scores: [0.49372475227573576, 0.45629093313936075,
0.4215764865653166, 0.4703336143410078, 0.4900214306483105, 0.46236955586870565]
Epoch 4 - Save Best Score: 0.4657 Model
Epoch 4 - Save Best Score: 0.4657 Model
Epoch 4 - Save Best Score: 0.4657 Model
----- Fold: 2 Result -----
----- Fold: 2 Result -----
----- Fold: 2 Result -----
Score: 0.4657 Scores: [0.49372475227573576, 0.45629093313936075,
0.4215764865653166, 0.4703336143410078, 0.4900214306483105, 0.46236955586870565]
Score: 0.4657 Scores: [0.49372475227573576, 0.45629093313936075,
0.4215764865653166, 0.4703336143410078, 0.4900214306483105, 0.46236955586870565]
Score: 0.4657 Scores: [0.49372475227573576, 0.45629093313936075,
0.4215764865653166, 0.4703336143410078, 0.4900214306483105, 0.46236955586870565]
----- Fold: 3 Training -----
----- Fold: 3 Training -----
----- Fold: 3 Training -----
DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",

```



```

"hidden_dropout": 0.0,
"hidden_dropout_prob": 0.0,
"hidden_size": 768,
"initializer_range": 0.02,
"intermediate_size": 3072,
"layer_norm_eps": 1e-07,
"max_position_embeddings": 512,
"max_relative_positions": -1,
"model_type": "deberta-v2",
"norm_rel_ebd": "layer_norm",
"num_attention_heads": 12,
"num_hidden_layers": 12,
"output_hidden_states": true,
"pad_token_id": 0,
"pooler_dropout": 0,
"pooler_hidden_act": "gelu",
"pooler_hidden_size": 768,
"pos_att_type": [
    "p2c",
    "c2p"
],
"position_biased_input": false,
"position_buckets": 256,
"relative_attention": true,
"share_att_key": true,
"transformers_version": "4.42.3",
"type_vocab_size": 0,
"vocab_size": 128100
}

```

```

DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",
  "hidden_dropout": 0.0,
  "hidden_dropout_prob": 0.0,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-07,
  "max_position_embeddings": 512,
  "max_relative_positions": -1,
  "model_type": "deberta-v2",
  "norm_rel_ebd": "layer_norm",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_hidden_states": true,

```

```

"pad_token_id": 0,
"pooler_dropout": 0,
"pooler_hidden_act": "gelu",
"pooler_hidden_size": 768,
"pos_att_type": [
    "p2c",
    "c2p"
],
"position_biased_input": false,
"position_buckets": 256,
"relative_attention": true,
"share_att_key": true,
"transformers_version": "4.42.3",
"type_vocab_size": 0,
"vocab_size": 128100
}

DebertaV2Config {
  "_name_or_path": "microsoft/deberta-v3-base",
  "attention_dropout": 0.0,
  "attention_probs_dropout_prob": 0.0,
  "hidden_act": "gelu",
  "hidden_dropout": 0.0,
  "hidden_dropout_prob": 0.0,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-07,
  "max_position_embeddings": 512,
  "max_relative_positions": -1,
  "model_type": "deberta-v2",
  "norm_rel_ebd": "layer_norm",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "output_hidden_states": true,
  "pad_token_id": 0,
  "pooler_dropout": 0,
  "pooler_hidden_act": "gelu",
  "pooler_hidden_size": 768,
  "pos_att_type": [
    "p2c",
    "c2p"
  ],
  "position_biased_input": false,
  "position_buckets": 256,
  "relative_attention": true,
  "share_att_key": true,
  "transformers_version": "4.42.3",

```

```
"type_vocab_size": 0,  
"vocab_size": 128100  
}
```

```
Epoch: [1][0/366] Elapsed 0m 1.3698627948760986s (remain 8m 19.999920129776s)  
Loss: 2.3001(2.3001) Grad: inf LR: 0.00002000  
Epoch: [1][20/366] Elapsed 0m 35.65700435638428s (remain 9m 45.7936429977417s)  
Loss: 0.1904(1.2432) Grad: 94009.1875 LR: 0.00001999  
Epoch: [1][40/366] Elapsed 1m 9.635453224182129s (remain 9m 11.988348728272967s)  
Loss: 0.1753(0.7305) Grad: 155063.3906 LR: 0.00001996  
Epoch: [1][60/366] Elapsed 1m 40.229840993881226s (remain 8m  
21.149204969406128s) Loss: 0.1939(0.5361) Grad: 66541.6875 LR: 0.00001991  
Epoch: [1][80/366] Elapsed 2m 14.537447929382324s (remain 7m  
53.372501973752605s) Loss: 0.1142(0.4418) Grad: 46436.6953 LR: 0.00001985  
Epoch: [1][100/366] Elapsed 2m 50.73058843612671s (remain 7m  
27.956494411619587s) Loss: 0.0913(0.3858) Grad: 87721.3516 LR: 0.00001977  
Epoch: [1][120/366] Elapsed 3m 22.85597538948059s (remain 6m 50.74143777208883s)  
Loss: 0.0934(0.3446) Grad: 78697.3594 LR: 0.00001967  
Epoch: [1][140/366] Elapsed 4m 6.012236595153809s (remain 6m 32.5727179709902s)  
Loss: 0.1122(0.3143) Grad: 69333.0078 LR: 0.00001955  
Epoch: [1][160/366] Elapsed 4m 44.9176504611969s (remain 6m 2.7833437549401197s)  
Loss: 0.0888(0.2920) Grad: 81031.0547 LR: 0.00001941  
Epoch: [1][180/366] Elapsed 5m 26.778427839279175s (remain 5m  
34.00005055395934s) Loss: 0.1315(0.2752) Grad: 39275.9766 LR: 0.00001926  
Epoch: [1][200/366] Elapsed 6m 0.43942737579345703s (remain 4m  
55.88311202490502s) Loss: 0.1610(0.2623) Grad: 101001.1016 LR: 0.00001909  
Epoch: [1][220/366] Elapsed 6m 37.54714751243591s (remain 4m  
20.834101309064295s) Loss: 0.1522(0.2496) Grad: 70961.3203 LR: 0.00001890  
Epoch: [1][240/366] Elapsed 7m 19.656426191329956s (remain 3m  
48.03756545193471s) Loss: 0.1055(0.2407) Grad: 44311.5977 LR: 0.00001870  
Epoch: [1][260/366] Elapsed 7m 54.32335138320923s (remain 3m  
10.819739062210601s) Loss: 0.0792(0.2319) Grad: 59317.4883 LR: 0.00001848  
Epoch: [1][280/366] Elapsed 8m 28.653918504714966s (remain 2m  
33.863284956942266s) Loss: 0.1566(0.2239) Grad: 82184.6641 LR: 0.00001824  
Epoch: [1][300/366] Elapsed 9m 3.1751551628112793s (remain 1m  
57.296960417218315s) Loss: 0.1631(0.2172) Grad: 50273.0391 LR: 0.00001799  
Epoch: [1][320/366] Elapsed 9m 37.47361660003662s (remain 1m  
20.954245317762116s) Loss: 0.1026(0.2111) Grad: 84002.5234 LR: 0.00001773  
Epoch: [1][340/366] Elapsed 10m 15.15145230293274s (remain 0m  
45.099080080860176s) Loss: 0.1289(0.2055) Grad: 41554.4844 LR: 0.00001745  
Epoch: [1][360/366] Elapsed 10m 43.85307002067566s (remain 0m  
8.917632548762867s) Loss: 0.0675(0.2002) Grad: 63603.7617 LR: 0.00001715  
Epoch: [1][365/366] Elapsed 10m 51.71582770347595s (remain 0m 0.0s) Loss:  
0.1340(0.1990) Grad: 54374.2656 LR: 0.00001708  
EVAL: [0/62] Elapsed 0m 1.9103620052337646s (remain 1m 56.532082319259644s)  
Loss: 0.1168(0.1168)  
EVAL: [20/62] Elapsed 0m 22.786691665649414s (remain 0m 44.48830277579172s)  
Loss: 0.0751(0.1098)
```

EVAL: [40/62] Elapsed 0m 44.86300563812256s (remain 0m 22.978612643916435s)  
 Loss: 0.1009(0.1132)  
 EVAL: [60/62] Elapsed 1m 6.4307966232299805s (remain 0m 1.0890294528398385s)  
 Loss: 0.0939(0.1122)  
 EVAL: [61/62] Elapsed 1m 6.515820741653442s (remain 0m 0.0s) Loss:  
 0.0769(0.1121)  
 Epoch 1 - avg\_train\_loss: 0.1990 avg\_val\_loss: 0.1121 time: 718s  
 Epoch 1 - avg\_train\_loss: 0.1990 avg\_val\_loss: 0.1121 time: 718s  
 Epoch 1 - avg\_train\_loss: 0.1990 avg\_val\_loss: 0.1121 time: 718s  
 Epoch 1 - Score: 0.4749 Scores: [0.5024306261044553, 0.48065709798763906,  
 0.4302532665191771, 0.457347128882137, 0.5169674434833702, 0.46150674299844385]  
 Epoch 1 - Score: 0.4749 Scores: [0.5024306261044553, 0.48065709798763906,  
 0.4302532665191771, 0.457347128882137, 0.5169674434833702, 0.46150674299844385]  
 Epoch 1 - Score: 0.4749 Scores: [0.5024306261044553, 0.48065709798763906,  
 0.4302532665191771, 0.457347128882137, 0.5169674434833702, 0.46150674299844385]  
 Epoch 1 - Save Best Score: 0.4749 Model  
 Epoch 1 - Save Best Score: 0.4749 Model  
 Epoch 1 - Save Best Score: 0.4749 Model  
 Epoch: [2][0/366] Elapsed 0m 1.1919691562652588s (remain 7m 15.068742036819458s)  
 Loss: 0.1727(0.1727) Grad: 303545.6250 LR: 0.00001706  
 Epoch: [2][60/366] Elapsed 1m 42.12880301475525s (remain 8m 30.644015073776245s)  
 Loss: 0.0923(0.1094) Grad: 167889.8750 LR: 0.00001610  
 Epoch: [2][80/366] Elapsed 2m 16.552065134048462s (remain 8m  
 0.46096991609647375s) Loss: 0.1194(0.1087) Grad: 144441.0156 LR: 0.00001575  
 Epoch: [2][100/366] Elapsed 2m 49.39075422286987s (remain 7m 24.44108781248042s)  
 Loss: 0.0842(0.1102) Grad: 69291.8047 LR: 0.00001540  
 Epoch: [2][120/366] Elapsed 3m 21.426045179367065s (remain 6m  
 47.84612453673503s) Loss: 0.0638(0.1104) Grad: 124949.9141 LR: 0.00001503  
 Epoch: [2][140/366] Elapsed 4m 0.46793079376220703s (remain 6m  
 23.725421479407828s) Loss: 0.1104(0.1100) Grad: 110344.1406 LR: 0.00001466  
 Epoch: [2][160/366] Elapsed 4m 35.01529884338379s (remain 5m  
 50.174759396855166s) Loss: 0.0736(0.1093) Grad: 114048.9453 LR: 0.00001427  
 Epoch: [2][180/366] Elapsed 5m 6.085354566574097s (remain 5m  
 12.849671794564642s) Loss: 0.1054(0.1084) Grad: 116303.8438 LR: 0.00001388  
 Epoch: [2][200/366] Elapsed 5m 41.22538113594055s (remain 4m  
 40.110387499652575s) Loss: 0.1066(0.1072) Grad: 127000.0469 LR: 0.00001348  
 Epoch: [2][220/366] Elapsed 6m 18.025073528289795s (remain 4m  
 8.025500731230863s) Loss: 0.0970(0.1080) Grad: 71429.5859 LR: 0.00001308  
 Epoch: [2][240/366] Elapsed 6m 50.715925455093384s (remain 3m  
 33.026932289986235s) Loss: 0.1724(0.1084) Grad: 207358.6250 LR: 0.00001267  
 Epoch: [2][260/366] Elapsed 7m 26.929975748062134s (remain 2m  
 59.799415530829606s) Loss: 0.1584(0.1087) Grad: 77780.5781 LR: 0.00001225  
 Epoch: [2][280/366] Elapsed 8m 10.074888467788696s (remain 2m  
 28.243293664633597s) Loss: 0.1133(0.1096) Grad: 79527.2500 LR: 0.00001183  
 Epoch: [2][300/366] Elapsed 8m 52.973628520965576s (remain 1m  
 55.093972936421096s) Loss: 0.1135(0.1089) Grad: 121753.8984 LR: 0.00001141  
 Epoch: [2][320/366] Elapsed 9m 22.443617343902588s (remain 1m  
 18.84723607624801s) Loss: 0.0865(0.1089) Grad: 103176.3594 LR: 0.00001098

Epoch: [2][340/366] Elapsed 9m 57.377596378326416s (remain 0m 43.79601146468667s) Loss: 0.1431(0.1085) Grad: 155213.2344 LR: 0.00001056  
 Epoch: [2][360/366] Elapsed 10m 34.3808491230011s (remain 0m 8.786438353504195s) Loss: 0.1580(0.1079) Grad: 107398.4531 LR: 0.00001013  
 Epoch: [2][365/366] Elapsed 10m 43.29887413978577s (remain 0m 0.0s) Loss: 0.0709(0.1076) Grad: 64407.2305 LR: 0.00001002  
 EVAL: [0/62] Elapsed 0m 1.8348383903503418s (remain 1m 51.92514181137085s) Loss: 0.1033(0.1033)  
 EVAL: [20/62] Elapsed 0m 22.619112491607666s (remain 0m 44.16112438837688s) Loss: 0.0695(0.1046)  
 EVAL: [40/62] Elapsed 0m 44.693995237350464s (remain 0m 22.89204634108195s) Loss: 0.1130(0.1075)  
 EVAL: [60/62] Elapsed 1m 6.260434627532959s (remain 0m 1.0862366332382436s) Loss: 0.0888(0.1080)  
 EVAL: [61/62] Elapsed 1m 6.345707416534424s (remain 0m 0.0s) Loss: 0.0588(0.1079)  
 Epoch 2 - avg\_train\_loss: 0.1076 avg\_val\_loss: 0.1079 time: 710s  
 Epoch 2 - avg\_train\_loss: 0.1076 avg\_val\_loss: 0.1079 time: 710s  
 Epoch 2 - avg\_train\_loss: 0.1076 avg\_val\_loss: 0.1079 time: 710s  
 Epoch 2 - Score: 0.4654 Scores: [0.5047564317852035, 0.4604429434210032, 0.42518514278992686, 0.4436699782665961, 0.4957893893627968, 0.46240736923857567]  
 Epoch 2 - Score: 0.4654 Scores: [0.5047564317852035, 0.4604429434210032, 0.42518514278992686, 0.4436699782665961, 0.4957893893627968, 0.46240736923857567]  
 Epoch 2 - Score: 0.4654 Scores: [0.5047564317852035, 0.4604429434210032, 0.42518514278992686, 0.4436699782665961, 0.4957893893627968, 0.46240736923857567]  
 Epoch 2 - Save Best Score: 0.4654 Model  
 Epoch 2 - Save Best Score: 0.4654 Model  
 Epoch 2 - Save Best Score: 0.4654 Model  
 Epoch: [3][0/366] Elapsed 0m 3.574061632156372s (remain 21m 44.532495737075806s) Loss: 0.0798(0.0798) Grad: 215456.4531 LR: 0.00001000  
 Epoch: [3][20/366] Elapsed 0m 37.397034883499146s (remain 10m 14.379858800343072s) Loss: 0.1321(0.1043) Grad: 147601.7188 LR: 0.00000957  
 Epoch: [3][40/366] Elapsed 1m 12.744885921478271s (remain 9m 36.63629084098625s) Loss: 0.1096(0.1041) Grad: 65632.1328 LR: 0.00000914  
 Epoch: [3][60/366] Elapsed 1m 45.227173805236816s (remain 8m 46.13586902618408s) Loss: 0.0819(0.1024) Grad: 85749.3750 LR: 0.00000872  
 Epoch: [3][80/366] Elapsed 2m 26.764885663986206s (remain 8m 36.39496807698856s) Loss: 0.0711(0.0994) Grad: 110455.9453 LR: 0.00000829  
 Epoch: [3][100/366] Elapsed 3m 6.8579418659210205s (remain 8m 10.270837569000719s) Loss: 0.1481(0.0996) Grad: 69011.0625 LR: 0.00000787  
 Epoch: [3][120/366] Elapsed 3m 42.51236820220947s (remain 7m 30.541571979680384s) Loss: 0.0911(0.0996) Grad: 70990.4844 LR: 0.00000746  
 Epoch: [3][140/366] Elapsed 4m 15.445570707321167s (remain 6m 47.62591070317205s) Loss: 0.1273(0.0995) Grad: 93225.9688 LR: 0.00000704  
 Epoch: [3][160/366] Elapsed 4m 47.28318691253662s (remain 6m 5.795362217826096s)

Loss: 0.0777(0.0994) Grad: 93784.7266 LR: 0.00000664  
 Epoch: [3][180/366] Elapsed 5m 22.155873775482178s (remain 5m 29.275340599249716s) Loss: 0.1093(0.0985) Grad: 44535.2891 LR: 0.00000624  
 Epoch: [3][200/366] Elapsed 6m 2.6157147884368896s (remain 4m 57.66961661737355s) Loss: 0.1231(0.0989) Grad: 114883.6250 LR: 0.00000584  
 Epoch: [3][220/366] Elapsed 6m 42.2312376499176s (remain 4m 23.907373118724195s) Loss: 0.0545(0.0988) Grad: 59576.7891 LR: 0.00000546  
 Epoch: [3][240/366] Elapsed 7m 16.021363258361816s (remain 3m 46.152159366370256s) Loss: 0.1039(0.0987) Grad: 96968.3125 LR: 0.00000508  
 Epoch: [3][260/366] Elapsed 7m 44.470677614212036s (remain 3m 6.856019729855348s) Loss: 0.0602(0.0983) Grad: 81584.5547 LR: 0.00000471  
 Epoch: [3][280/366] Elapsed 8m 27.174206018447876s (remain 2m 33.41568509454828s) Loss: 0.0913(0.0985) Grad: 113304.1562 LR: 0.00000435  
 Epoch: [3][300/366] Elapsed 9m 0.9588370323181152s (remain 1m 56.818353511962414s) Loss: 0.1210(0.0989) Grad: 156159.2188 LR: 0.00000400  
 Epoch: [3][320/366] Elapsed 9m 35.83564853668213s (remain 1m 20.724623626637708s) Loss: 0.0693(0.0987) Grad: 84955.6250 LR: 0.00000367  
 Epoch: [3][340/366] Elapsed 10m 9.791019201278687s (remain 0m 44.7060864517066s) Loss: 0.1453(0.0983) Grad: 169672.4062 LR: 0.00000334  
 Epoch: [3][360/366] Elapsed 10m 41.5546395778656s (remain 0m 8.885798332103377s) Loss: 0.0927(0.0981) Grad: 119223.5625 LR: 0.00000303  
 Epoch: [3][365/366] Elapsed 10m 48.77695083618164s (remain 0m 0.0s) Loss: 0.0727(0.0979) Grad: 62907.6250 LR: 0.00000295  
 EVAL: [0/62] Elapsed 0m 1.812776803970337s (remain 1m 50.57938504219055s) Loss: 0.1125(0.1125)  
 EVAL: [20/62] Elapsed 0m 22.60436511039734s (remain 0m 44.132331882204326s) Loss: 0.0649(0.1015)  
 EVAL: [40/62] Elapsed 0m 44.68093657493591s (remain 0m 22.885357757894013s) Loss: 0.1082(0.1029)  
 EVAL: [60/62] Elapsed 1m 6.247821807861328s (remain 0m 1.086029865702642s) Loss: 0.0863(0.1028)  
 EVAL: [61/62] Elapsed 1m 6.332708358764648s (remain 0m 0.0s) Loss: 0.0438(0.1027)  
 Epoch 3 - avg\_train\_loss: 0.0979 avg\_val\_loss: 0.1027 time: 715s  
 Epoch 3 - avg\_train\_loss: 0.0979 avg\_val\_loss: 0.1027 time: 715s  
 Epoch 3 - avg\_train\_loss: 0.0979 avg\_val\_loss: 0.1027 time: 715s  
 Epoch 3 - Score: 0.4541 Scores: [0.494218499586883, 0.44751973482161544, 0.41720802192955, 0.43740445474977185, 0.47866223588460993, 0.4495595162370406]  
 Epoch 3 - Score: 0.4541 Scores: [0.494218499586883, 0.44751973482161544, 0.41720802192955, 0.43740445474977185, 0.47866223588460993, 0.4495595162370406]  
 Epoch 3 - Score: 0.4541 Scores: [0.494218499586883, 0.44751973482161544, 0.41720802192955, 0.43740445474977185, 0.47866223588460993, 0.4495595162370406]  
 Epoch 3 - Save Best Score: 0.4541 Model  
 Epoch 3 - Save Best Score: 0.4541 Model  
 Epoch 3 - Save Best Score: 0.4541 Model  
 Epoch: [4][0/366] Elapsed 0m 1.2563281059265137s (remain 7m 38.55975866317749s) Loss: 0.0907(0.0907) Grad: 180172.9219 LR: 0.00000294  
 Epoch: [4][20/366] Elapsed 0m 38.58701968193054s (remain 10m

33.929609060287476s) Loss: 0.1077(0.0892) Grad: 176703.1406 LR: 0.00000264  
 Epoch: [4][40/366] Elapsed 1m 12.170364379882812s (remain 9m 32.08215666980277s)  
 Loss: 0.1089(0.0911) Grad: 194625.0625 LR: 0.00000236  
 Epoch: [4][60/366] Elapsed 1m 45.95904088020325s (remain 8m 49.795204401016235s)  
 Loss: 0.0925(0.0895) Grad: 110701.1406 LR: 0.00000209  
 Epoch: [4][80/366] Elapsed 2m 21.12161135673523s (remain 8m 16.53900292184619s)  
 Loss: 0.1124(0.0910) Grad: 140155.7812 LR: 0.00000183  
 Epoch: [4][100/366] Elapsed 2m 57.37642741203308s (remain 7m 45.39359667513634s)  
 Loss: 0.1090(0.0904) Grad: 158415.1562 LR: 0.00000159  
 Epoch: [4][120/366] Elapsed 3m 34.74260067939758s (remain 7m  
 14.809398069854637s) Loss: 0.0739(0.0907) Grad: 124588.8750 LR: 0.00000137  
 Epoch: [4][140/366] Elapsed 4m 10.073967933654785s (remain 6m  
 39.054204149449106s) Loss: 0.1358(0.0915) Grad: 40761.8047 LR: 0.00000116  
 Epoch: [4][160/366] Elapsed 4m 46.96531057357788s (remain 6m 5.390612842133351s)  
 Loss: 0.0781(0.0925) Grad: 63891.5234 LR: 0.00000097  
 Epoch: [4][180/366] Elapsed 5m 24.498814821243286s (remain 5m  
 31.670059347679626s) Loss: 0.0694(0.0917) Grad: 61247.5391 LR: 0.00000079  
 Epoch: [4][200/366] Elapsed 6m 0.13938164710998535s (remain 4m  
 55.636805829717105s) Loss: 0.0972(0.0927) Grad: 80241.3438 LR: 0.00000063  
 Epoch: [4][220/366] Elapsed 6m 37.34483790397644s (remain 4m  
 20.701364235640654s) Loss: 0.0957(0.0929) Grad: 51270.6719 LR: 0.00000049  
 Epoch: [4][240/366] Elapsed 7m 13.157158374786377s (remain 3m  
 44.66657592053241s) Loss: 0.0933(0.0924) Grad: 85066.0938 LR: 0.00000037  
 Epoch: [4][260/366] Elapsed 7m 46.19436502456665s (remain 3m 7.549457193791113s)  
 Loss: 0.0789(0.0917) Grad: 63073.6680 LR: 0.00000026  
 Epoch: [4][280/366] Elapsed 8m 22.505900621414185s (remain 2m  
 32.003564244911786s) Loss: 0.0727(0.0923) Grad: 52937.3398 LR: 0.00000017  
 Epoch: [4][300/366] Elapsed 8m 56.14234232902527s (remain 1m  
 55.778246682347685s) Loss: 0.1023(0.0927) Grad: 40710.5312 LR: 0.00000010  
 Epoch: [4][320/366] Elapsed 9m 34.90193223953247s (remain 1m 20.59372881862612s)  
 Loss: 0.0872(0.0923) Grad: 44105.8086 LR: 0.00000005  
 Epoch: [4][340/366] Elapsed 10m 11.699221849441528s (remain 0m  
 44.84598400655727s) Loss: 0.0896(0.0926) Grad: 96091.2188 LR: 0.00000002  
 Epoch: [4][360/366] Elapsed 10m 47.10202074050903s (remain 0m  
 8.962631866212064s) Loss: 0.0787(0.0924) Grad: 119826.6797 LR: 0.00000000  
 Epoch: [4][365/366] Elapsed 10m 55.02571129798889s (remain 0m 0.0s) Loss:  
 0.0588(0.0922) Grad: 49147.3945 LR: 0.00000000  
 EVAL: [0/62] Elapsed 0m 1.8129119873046875s (remain 1m 50.58763122558594s) Loss:  
 0.1075(0.1075)  
 EVAL: [20/62] Elapsed 0m 22.60424542427063s (remain 0m 44.13209820929028s) Loss:  
 0.0655(0.1007)  
 EVAL: [40/62] Elapsed 0m 44.68479657173157s (remain 0m 22.887334829423494s)  
 Loss: 0.1128(0.1023)  
 EVAL: [60/62] Elapsed 1m 6.256364107131958s (remain 0m 1.0861699033956s) Loss:  
 0.0850(0.1025)  
 EVAL: [61/62] Elapsed 1m 6.341490268707275s (remain 0m 0.0s) Loss:  
 0.0444(0.1024)  
 Epoch 4 - avg\_train\_loss: 0.0922 avg\_val\_loss: 0.1024 time: 721s

```

Epoch 4 - avg_train_loss: 0.0922  avg_val_loss: 0.1024  time: 721s
Epoch 4 - avg_train_loss: 0.0922  avg_val_loss: 0.1024  time: 721s
Epoch 4 - Score: 0.4533  Scores: [0.49273691873848935, 0.44564665171712375,
0.41667778234786473, 0.43766371780012686, 0.47829108626065014,
0.4487808732895121]
Epoch 4 - Score: 0.4533  Scores: [0.49273691873848935, 0.44564665171712375,
0.41667778234786473, 0.43766371780012686, 0.47829108626065014,
0.4487808732895121]
Epoch 4 - Score: 0.4533  Scores: [0.49273691873848935, 0.44564665171712375,
0.41667778234786473, 0.43766371780012686, 0.47829108626065014,
0.4487808732895121]
Epoch 4 - Save Best Score: 0.4533 Model
Epoch 4 - Save Best Score: 0.4533 Model
Epoch 4 - Save Best Score: 0.4533 Model
----- Fold: 3 Result -----
----- Fold: 3 Result -----
----- Fold: 3 Result -----
Score: 0.4533  Scores: [0.49273691873848935, 0.44564665171712375,
0.41667778234786473, 0.43766371780012686, 0.47829108626065014,
0.4487808732895121]
Score: 0.4533  Scores: [0.49273691873848935, 0.44564665171712375,
0.41667778234786473, 0.43766371780012686, 0.47829108626065014,
0.4487808732895121]
Score: 0.4533  Scores: [0.49273691873848935, 0.44564665171712375,
0.41667778234786473, 0.43766371780012686, 0.47829108626065014,
0.4487808732895121]
----- CV -----
----- CV -----
----- CV -----
Score: 0.4565  Scores: [0.4917565941909823, 0.44713635660179213,
0.41584525728156335, 0.4528075124023095, 0.47988471989412373,
0.4517777547898269]
Score: 0.4565  Scores: [0.4917565941909823, 0.44713635660179213,
0.41584525728156335, 0.4528075124023095, 0.47988471989412373,
0.4517777547898269]
Score: 0.4565  Scores: [0.4917565941909823, 0.44713635660179213,
0.41584525728156335, 0.4528075124023095, 0.47988471989412373,
0.4517777547898269]

```

[ ]: