**Embeddings experiments:**

**Embeddings\_processing\_2.py**

* No white
* With split (hash)
* Max pooling
* Not balanced
* Bag size = 5

**Image\_embeddings\_mil\_2.csv**

* Rows 11785
* Bags 0 = 447, Bags 1 = 11337

**Ludwig experiment**

* Resnet\_h\_mil\_config.yml
* Split: fixed
* Training: 8798; validation: 809; test: 2177
* │ │ train │ validation │ test │
* ╞═══════════════╪═════════╪══════════════╪════════╡
* │ accuracy │ 0.9531 │ 0.9963 │ 0.9761 │
* ├───────────────┼─────────┼──────────────┼────────┤
* │ loss │ 0.3271 │ 0.0198 │ 0.1475 │
* ├───────────────┼─────────┼──────────────┼────────┤
* │ precision │ 0.9555 │ 0.9963 │ 0.9761 │
* ├───────────────┼─────────┼──────────────┼────────┤
* │ recall │ 0.9973 │ 1.0000 │ 1.0000 │
* ├───────────────┼─────────┼──────────────┼────────┤
* │ roc\_auc │ 0.6243 │ 0.9177 │ 0.6154 │
* ├───────────────┼─────────┼──────────────┼────────┤
* │ specificity │ 0.0051 │ 0.0000 │ 0.0000 │
* ├───────────────┼─────────┼──────────────┼────────┤
* │ combined\_loss │ 0.3271 │ 0.0198 │ 0.1475 │

**Embeddings\_processing\_3.py**

* No white
* With split (hash)
* Max pooling
* Balanced (remove excess)
* Bag\_size = 5

**Image\_embeddings\_mil\_3.csv**

* rows 903
* bags 0 = 451, bags 1 = 451

**Ludwig experiment**

* resnet\_h\_mil\_config.yml
* split: fixed
* Training: 784; Validation: 2; Test 116

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.5587 │ 1.0000 │ 0.6207 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.9504 │ 0.1647 │ 0.7768 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.5578 │ 1.0000 │ 0.6029 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.5663 │ 1.0000 │ 0.7069 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.5781 │ 1.0000 │ 0.6204 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.5510 │ 1.0000 │ 0.5345 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.9504 │ 0.1647 │ 0.7768 │**

**Embeddings\_processing\_4.py**

* No white
* With split (hash)
* Max pooling
* Balanced (remove excess)
* Bag\_size = 3 – 7

**Image\_embeddings\_mil\_4.csv**

* rows 1357
* bags 0 = 678, bags 1 = 678

**Ludwig experiment**

* resnet\_h\_mil\_config.yml
* split: fixed
* Training: 1128; Validation: 12; Test 216

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.7332 │ 1.0000 │ 0.6713 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.5796 │ 0.1467 │ 0.7333 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.7369 │ 1.0000 │ 0.6529 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.7252 │ 1.0000 │ 0.7315 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.8116 │ 1.0000 │ 0.7504 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.7411 │ 1.0000 │ 0.6111 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.5796 │ 0.1467 │ 0.7333**

**Image\_embeddings\_mil\_5.csv**

* rows 1357
* bags 0 = 678, bags 1 = 678
* no validation in split

**Ludwig experiment**

* resnet\_h\_mil\_config.yml
* split: fixed
* Training: 1128; Test 228

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.9991 │ │ 0.6491 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.0031 │ │ 1.9118 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 1.0000 │ │ 0.6214 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.9982 │ │ 0.7632 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 1.0000 │ │ 0.7073 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 1.0000 │ │ 0.5351 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.0031 │ │ 1.9118 │**

**Embeddings\_processing\_6.py**

* All data
* With split (hash)
* Max pooling
* Balanced (remove excess)
* Bag\_size = 3 – 7

**Image\_embeddings\_mil\_6.csv**

* rows 6927
* bags 0 = 3463, bags 1 = 3463

**Ludwig experiment**

* resnet\_h\_mil\_config.yml
* split: fixed
* Training: 5538; Validation: 658; Test 730

**│ │ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.7028 │ 0.6413 │ 0.6726 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.6310 │ 0.8187 │ 0.7332 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.7055 │ 0.6099 │ 0.6352 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.6963 │ 0.7842 │ 0.8110 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.7751 │ 0.6846 │ 0.7521 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.7093 │ 0.4985 │ 0.5342 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.6310 │ 0.8187 │ 0.7332 │**

**Embeddings\_processing\_7.py**

* All data
* No split
* No merge
* Only embeddings in rows
* No bags

**Image\_embeddings\_mil\_7.csv**

* rows 289,879
* tiles labeled 0 = 152151, tiles labeled 1 = 152151

**Ludwig experiment\_7\_a**

* resnet\_h\_mil\_config.yml
* split: hash
* Training: 209763; Validation: 19704; Test 60411

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.7514 │ 0.6062 │ 0.5847 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.5333 │ 0.8920 │ 0.9619 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.7394 │ 0.5451 │ 0.6087 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.7328 │ 0.4660 │ 0.4641 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.8359 │ 0.6517 │ 0.6319 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.7680 │ 0.7105 │ 0.7043 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.5333 │ 0.8920 │ 0.9619 │**

**Ludwig experiment\_7\_b**

* resnet\_h\_mil\_config.yml
* split: split: random [7,1,2]
* Training: 202915; Validation: 28987; Test 57976

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.8276 │ 0.7778 │ 0.7724 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.4905 │ 0.7406 │ 0.7468 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.8231 │ 0.7696 │ 0.7636 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.8126 │ 0.7557 │ 0.7521 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.9109 │ 0.8605 │ 0.8578 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.8413 │ 0.7975 │ 0.7907 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.4905 │ 0.7406 │ 0.7468 │**

**Embeddings\_processing\_8.py**

* No white
* No split
* No merge
* Only embeddings in rows
* No bags

**Image\_embeddings\_mil\_8.csv**

* rows 58,919
* tiles labeled 0 = 30471, tiles labeled 1 = 28447

**Ludwig experiment\_8\_a**

* resnet\_h\_mil\_config.yml
* split: hash
* Training: 44103; Validation: 3726; Test 11089

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.8897 │ 0.7622 │ 0.6166 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.3565 │ 1.1813 │ 1.9460 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.8877 │ 0.8443 │ 0.5828 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.8793 │ 0.7732 │ 0.5729 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.9569 │ 0.8288 │ 0.6509 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.8992 │ 0.7423 │ 0.6534 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.3565 │ 1.1813 │ 1.9460 │**

**Ludwig experiment\_8\_b**

* resnet\_h\_mil\_config.yml
* split: random [7,1,2]
* Training: 41243; Validation: 5891; Test 11784

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.9445 │ 0.8465 │ 0.8435 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.2157 │ 0.9224 │ 0.9145 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.9425 │ 0.8376 │ 0.8275 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.9432 │ 0.8424 │ 0.8468 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.9862 │ 0.9208 │ 0.9208 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.9457 │ 0.8504 │ 0.8406 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.2157 │ 0.9224 │ 0.9145 │**

**Embeddings\_processing\_9.py**

* All data
* Random
* Max pooling
* No balance
* Bag\_size = 3 – 7

**Image\_embeddings\_mil\_9.csv**

* rows 57,972
* tiles labeled 0 = 3373, tiles labeled 1 = 54598

**Ludwig experiment**

* resnet\_h\_mil\_config\_balanced.yml
* split: random [7,1,2]
* Training: 40580; Validation: 5797; Test 11594

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.9409 │ 0.9427 │ 0.9426 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.3510 │ 0.3168 │ 0.3208 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.9415 │ 0.9427 │ 0.9426 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.9994 │ 1.0000 │ 1.0000 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.6964 │ 0.7432 │ 0.7490 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.0004 │ 0.0000 │ 0.0000 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.3510 │ 0.3168 │ 0.3208 │**

**Embeddings\_processing\_10.py**

* No white
* Random
* Max pooling
* No balance
* Bag\_size = 3 – 7

**Image\_embeddings\_mil\_10.csv**

* rows 11789
* tiles labeled 0 = 636, tiles labeled 1 = 11152

**Ludwig experiment**

* resnet\_h\_mil\_config\_balanced.yml
* split: random [7,1,2]
* Training: 8252; Validation: 1178; Test 2358

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.9596 │ 0.9508 │ 0.9432 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.1671 │ 0.3248 │ 0.3242 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.9594 │ 0.9514 │ 0.9456 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.9996 │ 0.9991 │ 0.9973 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.9367 │ 0.7933 │ 0.8327 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.2634 │ 0.0500 │ 0.0000 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.1671 │ 0.3248 │ 0.3242 │**

**Embeddings\_processing\_11.py**

* No White
* Random
* Max pooling
* Balanced (Turns)
* Bag\_size = 7

**Image\_embeddings\_mil\_11.csv**

* rows 4690
* tiles labeled 0 = 2345, tiles labeled 1 = 2344

**Ludwig experiment**

* resnet\_h\_mil\_config\_balanced.yml
* split: random [7,1,2]
* Training: 3283; Validation: 468; Test 938

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.6208 │ 0.5470 │ 0.5480 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.6809 │ 0.8879 │ 0.8654 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.6213 │ 0.5816 │ 0.6337 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.6239 │ 0.2500 │ 0.2319 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.6684 │ 0.5371 │ 0.5742 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.6176 │ 0.8292 │ 0.8654 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.6809 │ 0.8879 │ 0.8654 │**

**Embeddings\_processing\_11.py**

* All data
* Random
* Max pooling
* Balanced (Turns)
* Bag\_size = 7

**Image\_embeddings\_mil\_12.csv**

* rows 23410
* tiles labeled 0 = 11705, tiles labeled 1 = 11704

**Ludwig experiment**

* resnet\_h\_mil\_config\_balanced.yml
* split: random [7,1,2]
* Training: 16387; Validation: 2340; Test 4682

**│ train │ validation │ test │**

**╞═══════════════╪═════════╪══════════════╪════════╡**

**│ accuracy │ 0.5777 │ 0.5449 │ 0.5369 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ loss │ 0.6791 │ 0.7006 │ 0.7052 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ precision │ 0.5768 │ 0.5425 │ 0.5295 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ recall │ 0.5742 │ 0.6661 │ 0.6749 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ roc\_auc │ 0.6132 │ 0.5615 │ 0.5592 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ specificity │ 0.5811 │ 0.4196 │ 0.3986 │**

**├───────────────┼─────────┼──────────────┼────────┤**

**│ combined\_loss │ 0.6791 │ 0.7006 │ 0.7052 │**

**Embeddings\_processing\_12.py**

* All data/no\_white
* Hash
* Max pooling
* Balanced (Turns)
* Bag\_size = 3-7

**Image\_embeddings\_mil\_13\_all\_data.csv**

* rows 57864
* tiles labeled 0 = 20228, tiles labeled 1 = 37635

**Image\_embeddings\_mil\_13\_no\_white.csv**

* rows 11789
* tiles labeled 0 = 4070, tiles labeled 1 = 7718

**Ludwig Experiment (All)**

* resnet\_h\_mil\_config.yml
* split: fixed – column split
* Training: 41300; validation: 5073; Test: 11490

│ train │ validation │ test │

╞═══════════════╪═════════╪══════════════╪════════╡

│ accuracy │ 0.7213 │ 0.6393 │ 0.7134 │

├───────────────┼─────────┼──────────────┼────────┤

│ loss │ 0.5880 │ 0.8359 │ 0.6804 │

├───────────────┼─────────┼──────────────┼────────┤

│ precision │ 0.7633 │ 0.6470 │ 0.7291 │

├───────────────┼─────────┼──────────────┼────────┤

│ recall │ 0.8168 │ 0.9428 │ 0.9415 │

├───────────────┼─────────┼──────────────┼────────┤

│ roc\_auc │ 0.7718 │ 0.6197 │ 0.6731 │

├───────────────┼─────────┼──────────────┼────────┤

│ specificity │ 0.5527 │ 0.1197 │ 0.1761 │

├───────────────┼─────────┼──────────────┼────────┤

│ combined\_loss │ 0.5880 │ 0.8359 │ 0.6804 │

╘═══════════════╧═════════╧══════════════╧════════╛

**Ludwig Experiment (Now)**

* resnet\_h\_mil\_config.yml
* split: fixed – column split
* Training: 8791; validation: 818; Test: 2179

│ │ train │ validation │ test │

╞═══════════════╪═════════╪══════════════╪════════╡

│ accuracy │ 0.7788 │ 0.8619 │ 0.6856 │

├───────────────┼─────────┼──────────────┼────────┤

│ loss │ 0.5490 │ 0.3717 │ 1.0319 │

├───────────────┼─────────┼──────────────┼────────┤

│ precision │ 0.8190 │ 0.8974 │ 0.7005 │

├───────────────┼─────────┼──────────────┼────────┤

│ recall │ 0.8384 │ 0.9344 │ 0.9251 │

├───────────────┼─────────┼──────────────┼────────┤

│ roc\_auc │ 0.8510 │ 0.8847 │ 0.6503 │

├───────────────┼─────────┼──────────────┼────────┤

│ specificity │ 0.6737 │ 0.5706 │ 0.2033 │

├───────────────┼─────────┼──────────────┼────────┤

│ combined\_loss │ 0.5490 │ 0.3717 │ 1.0319 │