|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Datasets | Missing rate | Training-Test  Image Text | CLIP  ViT-B/32  Sum Average | CLIP  RES50X16  Sum Average | ViLT | CLIP  RES50X16 |
| MM-IMDb  (F1-Macro) | 70% | 100% 30%  30% 100%  65% 65% | 44.40  52.53  50.34 49.99 | 53.22  52.41  54.87 54.28 | 35.13  37.73  36.26 | 52.71  55.63  52.07 |
| Food101  (Accuracy) | 70% | 100% 30%  30% 100%  65% 65% | 64.06  84.82  78.69 78.63 | 71.87  85.49  81.63 81.60 | 66.29  76.66  69.25 | 80.27  87.40  82.81 |
| Hateful Memes  (AUROC) | 70% | 100% 30%  30% 100%  65% 65% | 65.28  67.51  65.45 65.47 | 71.55  65.38  67.29 67.26 | 60.78  61.64  62.58 | 70.44  66.40  65.43 |

**CLIP-FIRST EXPERIMENTS**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Datasets | Missing rate | Fusion | Training-Test  Image Text | CLIP  RES50X16 | Missing rate | Fusion | CLIP  RES50X16 |
| MM-IMDb  (F1-Macro) | 70% | Sum | 100% 30%  30% 100%  65% 65% | 49.19  53.36  52.51 | 0% | Sum | **61.26** |
| 70% | Average | 100% 30%  30% 100%  65% 65% | 51.19  53.45  54.44 | 0% | Average | **60.95** |
| Food101  (Accuracy) | 70% | Sum | 100% 30%  30% 100%  65% 65% | 78.91  86.26  84.11 | 0% | Sum | **94.29** |
| 70% | Average | 100% 30%  30% 100%  65% 65% | 78.76  86.50  84.00 | 0% | Average | **91.39** |
| Hateful Memes  (AUROC) | 70% | Sum | 100% 30%  30% 100%  65% 65% | 72.54  66.07  65.68 | 0% | Sum | **76.27** |
| 70% | Average | 100% 30%  30% 100%  65% 65% | 72.71  66.16  65.72 | 0% | Average | **75.27** |

**CLIP**

**OPENCLIP-RN50-CC12M**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Food101  (Accuracy) | 70% | Sum | 100% 30%  30% 100%  65% 65% | 58.54  83.90  72.52 | 0% | Sum | **87.65** |
| 70% | Average | 100% 30%  30% 100%  65% 65% | 58.56  83.83  71.87 | 0% | Average | **87.51** |

**CLIP**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Food101  (Accuracy) | 70% | Sum | 100% 30%  30% 100%  65% 65% | 78.91  86.26  84.11 | 0% | Sum | **94.29** |
| 70% | Average | 100% 30%  30% 100%  65% 65% | 78.76  86.50  84.00 | 0% | Average | **91.39** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Datasets | Missing rate | Fusion | Training-Test  Image Text | CLIP  RES50X16 | Missing rate | Fusion | CLIP  RES50X16 |
| MM-IMDb  (F1-Macro) | 70% | Sum | 100% 30%  30% 100%  65% 65% | 32.89  34.83  36.26 | 0% | Sum | **45.88** |
| 70% | Average | 100% 30%  30% 100%  65% 65% | 32.78  35.50  36.13 | 0% | Average | **44.90** |
| Food101  (Accuracy) | 70% | Sum | 100% 30%  30% 100%  65% 65% | --  --  -- | 0% | Sum | **--** |
| 70% | Average | 100% 30%  30% 100%  65% 65% | --  --  -- | 0% | Average | **--** |
| Hateful Memes  (AUROC) | 70% | Sum | 100% 30%  30% 100%  65% 65% | 57.23  57.57  59.52 | 0% | Sum | **65.76** |
| 70% | Average | 100% 30%  30% 100%  65% 65% | 56.90  57.55  59.68 | 0% | Average | **65.29** |

**OPENCLIP-RN50-CC12M**