

X-ray Images

Caption

RNN

CNN

Visual Features

X-ray Reports

Word Embeddings

| **References** | **Data (Images, reports)** | **Architecture** | **BLEU-1** | **BLEU-2** | **BLEU-3** | **BLEU-4** | **METEOR** | **ROUGE** | **CIDEr** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Shin et al. 2016 | OpenI (7470, 3955) | CNN–RNN | 0.972 | 0.671 | 0.149 | 0.028 | – | – | – |
| Zhang et al. 2017 | Bladder Cancer (1000, 5000) | CNN–LSTM–ATT | 0.912 | 0.829 | 0.75 | 0.677 | 0.396 | 0.701 | 0.0204 |
| Jing et al.2017 | IU X-Ray (7470, 7470) | CNN–HLSTM–ATT | 0.517 | 0.386 | 0.306 | 0.247 | 0.217 | 0.447 | 0.327 |
| Wang et al. 2018 | ChestX-ray14 (−, −) | CNN–LSTM–ATT | 0.2860 | 0.1597 | 0.1038 | 0.0736 | 0.1076 | 0.2263 | – |
| Xue et al. 2018 | IU X-Ray (7470, 7470) | Recurrent CNN–LSTM–ATT | 0.464 | 0.358 | 0.270 | 0.195 | 0.274 | 0.366 | – |
| Han et al. 2018 | Lumbar Spinal MRI (253, 253) | Weakly Supervised CNN–LSTM | – | – | – | – | – | – | – |
| Tian et al. 2018 | CT (−, −) | CNN–LSTM | – | – | – | 0.766 | – | – | – |
| Zeng et al. 2018 | Ultrasound Image (−, −) | CNN–LSTM | 0.22 | 0.13 | 0.09 | - | 0.10 | 0.39 | 0.90 |
| Ma et al. 2018 | Pathology (−, −) | CNN–LSTM | – | – | – | – | – | – | – |
| Harzig et al. 2019 | IU X-Ray (7470, 3955) | CNN–HLSTM–DualLSTM–ATT | 0.373 | 0.246 | 0.175 | 0.126 | 0.163 | 0.315 | 0.359 |
| Yuan et al. 2019 | CheXpert (6248, -) | Muti-view CNN–LSTM–ATT–Medical Concepts | 0.529 | 0.372 | 0.315 | 0.255 | 0.343 | 0.453 | - |
| Lee et al. 2019 | DDSM FFDM2.0 (605, 605) | CNN–LSTM–ATT | 0.4070 | 0.2296 | 0.1354 | 0.0871 | - | 0.2650 | 0.1366 |
| Liu et al. 2019 | MIMIC-CXR(327,281, 141,783) | CNN–HLSTM–RL | 0.313 | 0.206 | 0.146 | 0.103 | 0.146 | 0.306 | 1.046 |
| Jing et al. 2019 | CX-CHR (−, −) | CMAS–RL | 0.428 | 0.361 | 0.323 | 0.290 | – | 0.504 | 2.968 |
| Gale et al. 2019 | Frontal Pelvic X-rays (50,363, -) | CNN–LSTM–ATT | 0.919 | 0.838 | 0.761 | 0.677 | – | – | – |
| Hasan et al. 2019 | Biomedical Images(164,614, -) | CNN–LSTM | 0.3211 | – | – | – | – | – | – |
| Sun et al. 2019 | INbreast (−, −) | CNN–LSTM | – | – | – | – | – | – | – |
| Xie et al. 2019 | – | CNN–LSTM–ATT | – | – | – | – | – | – | – |
| Li et al. 2019 | IU X-Ray (7470, 7470) | CNN–LSTM–ATT | 0.419 | 0.280 | 0.201 | 0.150 | - | 0.371 | 0.553 |
| Yin et al. 2020 | Two image-paragraph pair data sets | Hierarchical RNN | – | – | – | – | – | – | – |
| Pino et al. 2020 | IU X-Ray (7470, 7470) | CNN–LSTM–ATT | 0.361 | 0.226 | 0.152 | 0.106 | - | 0.314 | 0.187 |
| Zeng et al. 2020 | Ultrasound image | CNN–LSTM | – | – | – | – | – | – | – |
| Xu et al. 2020 | IU X-Ray (7470, 7470) and MIMIC-CXR | Reinforce CNN–LSTM | 0.412 | 0.279 | 0.206 | 0.157 | 0.179 | 0.342 | 0.411 |
| Singh et al. 2021 | IU X-Ray (−, −) | CNN–LSTM- | 23.07 | 11.86 | 7.05 | 4.75 | 11.11 | 23.15 | 19.78 |
| Yang et al. 2021) | Ultrasound image | Adaptive Multimo-dal ATT | – | – | – | – | – | – | – |
| Najdenkoska et al. 2021 | IU X-Ray (7470, 7470) and MIMIC-CXR | CNN–LSTM–ATT | – | – | – | – | – | – | – |
| Oa et al. 2021 | IU X-Ray (7470, 7470) | Condition GPT2 | 0.387 | 0.245 | 0.166 | 0.111 | 0.164 | 0.289 | 0.257 |
| Liu et al. 2021 | COVID-19 cases (1104, 368) | Medical visual language BERT | – | – | – | – | – | – | – |
| Han et al. 2021 | spinal image data set | Neural-symbolic learning (NSL) framework | – | – | – | – | – | – | – |
| Wu et al. 2022 | skin pathological image data set (1147, 1147) | CNN–LSTM–ATT | – | – | – | – | – | – | – |
| Chang et al. 2022 | lung CT scans (458, 458) |  | – | – | – | – | – | – | – |

Steps in the data processing: Problem is the image is frontal or lateral