|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reference** | **Model** | **Dataset** | **Significant Factor** | **Evaluation Measures** |
| (Xie et al., 2024) | TMSON(Trust Worthy multimodal sentimental ordinal Network) | CMU-MOSI , CMU-MOSEI and SIMS | Bayesian Fusion to obtain multimodal distributions from unimodal distributions. | Average F1-score : 80 |
| (Xiao et al., 2022) | Multimodal sentimental analysis | CMU-MOSI , CMU-MOSEI and YouTube | Multi – channel attentive graph convolutional network of audio, video and text . | Accuracy:80.6  F1-score:80.3 |
| (Z. Zhao et al., 2023) | Multimodal Emotion Recognition | Interactive Emotional Dyadic Motion Capture(IEMOCAP) corpus. | Transformer based deep scale fusion of text and audio. | 82.08% WA and 82.57% UA |
| (Yuan et al., 2024) | Multimodal sentimental Analysis | CMU-MOSEI and SIMS. | incorporates consistency-based pseudo-label technique. | F1-Score:84.3 |
| (Zhu et al., 2022) | Multimodal Sentiment analysis. | B-T4SA and IMDb movie reviews datasets | image-text matching and  self-supervised learning | accuracy of 84.2% |
| (Zhang et al., 2024) | Multimodal Sarcasm and sentimental analysis. | Memotion and MUStARD | Multimodal multitask interaction learning framework for joint detection of sarcasm and  Sentiment. | F1- Score: 61.37 |
| (X. Zhao et al., 2022) | Multimodal Sentimental Analysis | CMU-MOSI and CMU-MOSEI | Each modality is reinforced by all other modalities via crossmodal attention at first, and then all modalities are fused via MAG | accuracy of  86.30% |
| (Liu et al., 2024) | Multimodal Sarcasm, sentiment and Emotion analysis | MUStARD and MELD. | Quantum Probability (QP) into the joint multi-modal analysis of sarcasm, sentiment, and emotion. | micro-F1 scores  of 83.7%, 74.89%, 37.69% |
| (Q. Wang et al., 2023) | Multimodal aspect based sentimental analysis | Twitter2015 and Twitter2017. | end-to-end Multimodal Aspect-Based Sentiment Analysis (MABSA) framework | Accuracy of 79.48 and F1-score of 75.11 |
| (Al-Tameemi et al., 2023) | Multimodal Sentimental classification | BG, Emotion-Getty (EMO-G) and Twitter dataset. | multi-view feature extraction and attentive interaction learning between visual and textual modalities. | accuracy and an F1-score  of 99.801%. |
| (Huang et al., 2024) | Multimodal Sentimental Analysis | CMU-MOSI and CMU-MOSEI | SIMSUF effectively fuses multimodal data using dominant modality guidance. | F1 – score of 86.12 |
| (Lin & Hu, 2024) | Multimodal Sentimental Analysis | CMU-MOSI and CMU-MOSEI | Multi-stage fusion dynamically fine-tunes multimodal representations using hybrid-modal attention. | F1 – score of 85.6 |
| (F. Qi et al., 2024) | Zero shot emotion recognition | LIRIS, YouTube-8 and YouTube-24, StockEmotion. | Zero-shot MER framework refines emotion embeddings using affective graph space | ZSL of 65.84 and Harmonic mean of 36.41 |
| (Tang et al., 2023) | Multimodal sentimental analysis | CMU-MOSI and CMU-MOSEI | Bi-directional attention block captures fine-grained multimodal sentiment via dynamic routing. | Accuracy and F1- score of 89.1 |
| (Y. Wang et al., 2022) | Multimodal video level sentimental analysis | CMU-MOSI and CMU-MOSEI | reduces unimodal representation differences, improving multimodal video sentiment analysis performance. | F1 – score of 87. 5 an accuracy of 84.3 |
| (Low et al., 2023) | Multimodal emotion recognition | IEMOCAP, CMU-MOSEI | universal single-source adversarial  perturbations framework | Accuracy of 78.12 |
| (Chen et al., 2022) | Multimodal sentimental analysis | CMU-MOSI and CMU-MOSEI | weighted cross-modal attention mechanism captures the temporal correlation information and the  spatial dependence information of each modality | Accuracy of 3.82 |
| (Xue et al., 2023) | Multimodal sentimental analysis | MVSA-Single, MVSA Multi and Multi-ZOL | filter noise before multimodal fusion | Accuracy of 78.34 and F1-score of 77.92 |
| (H. Sun et al., 2023) | Multimodal sentimental analysis and depression detection. | CMU-MOSI ,CMU-MOSEI and AVEC2019 | generate  the corresponding interacted features by calculating source-target attention | Accuracy of 86.9 and F1-score of 86.8 |
| (He & Hu, 2022) | Multimodal sentimental analysis | CMU-MOSI and CMU-MOSEI | explore the time-dependent interactions  among different modalities. | Accuracy of 85.6 |
| (Fang et al., 2024) | Multimodal sentimental analysis | CMU-MOSI and CMU-MOSEI | Coarse-grained  Interaction Network (CIN) exploits the unique characteristics  of different modalities at a coarse-grained level | Accuracy of 85.76 |
| (Jia et al., 2024) | Multimodal sentimental classification | TWITTER-15 and TWITTER-17 | interaction of image, text, and  targets along the modal-axis, sequential-axis, and feature channel axis | Accuracy of 79.94 and F1-score of 79.66 |
| (X. Zhao et al., 2024) | Multimodal sentimental analysis | CMU-MOSI and CMU-MOSEI | MCER model reduces redundancy, enhances multimodal interaction, and filters unimodal noise effectively. | Accuracy of 86.73 |
| (Q. Qi et al., 2022) | Multimodal sentimental analysis | CMU-MOSI and CMU-MOSEI | encoding-decoding network solves long-term dependencies and modality weight | F1-score of 81.82 |
| (Katsurai & Ichi Satoh, n.d.) | Image sentimental analysis | Flickr and Instagram datasets. | latent correlations among multiple views constructed using SentiWordNe.t | Accuracy of 74.77 ± 0.82% |
| (Jain et al., 2024) | Multimodal sentimental analysis | CMU-MOSI | Co-learning fosters model explainability in multimodal sentiment analysis through modality dominance insights. | Accuracy of 73.03 and F1-score of 75 |
| (Qian et al., 2024) | Multimodal sentimental analysis | CMU-MOSI and CMU-MOSEI | Capsule networks uncover deep sentiment structures and improve cross-modal interaction understanding. | Accuracy of 84.8 and F1-score of 84.7 |
| (H. Zhao et al., 2024) | Multimodal sentimental analysis | *MULTI-ZOL TWITTER-15 AND TWITTER-17 , MASAD DATASETs* | Survey on Multimodal Aspect-Based Sentiment Analysis (MABSA): concepts,methods, evaluations, and future research trends | Accuracy of 78.6 and F1-score of 74.19 |
| (X. Sun et al., 2024) | Multimodal sentimental analysis | CMU-MOSI and CMU-MOSEI | uses gated fusion and multi-task learning. | Accuracy of 86% and F1-score of 85.8 |
| (He et al., 2021) | Multimodal sentimental analysis | CMU-MOSEI | Time squeeze fusion with unimodal reinforced Transformer improves multimodal sentiment analysis performance. | Accuracy of 82.2 and F1-score of 82.4 |
| (Huan et al., 2024) | Multimodal sentimental analysis | CMU-MOSI, CMU-MOSEI , MELD and UR-FUNNY datasets. | UniMF tackles missing and unaligned multimodal sequences with transformers effectively. | Accuracy of 82% |