|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reference** | **Model** | **Dataset** | **Significant Factor** | **Evaluation Measures** |
| Chorowski, J., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y. (2015) | Attention-Based Recurrent Sequence Generators (ARSG) | TIMIT corpus | Incorporation of convolutional features for location-awareness. | Phoneme Error Rate (PER):  Baseline model: 18.7% (test set). |
| Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., & An, W. (2018) | Faster R-CNN (Region-based Convolutional Neural Network) | A custom dataset containing over 100,000 image frames from far-field surveillance videos at 25 construction sites collected over one year | High precision and recall rates in detecting non-hardhat use (NHU).  Adaptability to various visual conditions (weather, illumination, occlusions, worker posture). | Precision: 95.7%  Recall: 94.9% |
| Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014) | RNN Encoder-Decoder | English-French translation task from the WMT’14 dataset | Use of adaptive gating mechanisms to improve memory capacity and training efficiency. | BLEU Scores:  Baseline: 33.3  Baseline + RNN Encoder-Decoder: 33.87  Baseline + CSLM + RNN Encoder-Decoder: 34.64 |
| Ben-Alon, L., & Sacks, R. (2017) | Agent-Based Simulation (ABS) | Data collected from field interviews and observations with:  13 superintendents and trade crew leaders across four residential tower projects.  Detailed parameters such as production rates, material supply, labor assignments, and decision-making patterns. | Determines task prioritization based on perceived profitability. Influence of material and design information flows on production control. | project duration (e.g., 156–232 days across scenarios), time distribution for crew activities (e.g., 46% work with design info, 16% waiting, 9% rework in Scenario 9), and rework rates reflecting workflow disruptions |
| Waehrer, G. M., Dong, X. S., Miller, T., Haile, E., & Men, Y. (2007) | Cost Model combining direct, indirect, and quality-of-life costs for occupational injuries. | 2002 data from the Bureau of Labor Statistics (BLS) on occupational injuries and illnesses and Census of Fatal Occupational Injuries (CFOI). | Fatal and nonfatal injuries.  Direct costs (medical expenses), indirect costs (wage and household productivity losses), and quality-of-life costs. | Average Fatal Injury Cost: $4 million.  Average Nonfatal Injury Cost: $42,000 per case.  Construction accounted for 15% of private industry injury costs |
| Liu, Y., Yang, G., Qiao, S., Liu, M., Qu, L., Han, N., Wu, T., Yuan, G., & Peng, Y. (2023) | Transfer Learning Classifier (TLC) | CIFAR10, CIFAR100 (synthetically imbalanced), Caltech101, and HAM10000 datasets. | Dynamically adjusts class distributions using F1-scores.  Enhances generalization and reduces overfitting | Classification Accuracy: TLC+COT achieved 85.96% on CIFAR10 (imbalance ratio 10) and 62.05% on CIFAR100 (imbalance ratio 100). |
| Kong, T., Fang, W., Love, P. E. D., Luo, H., Xu, S., & Li, H. (2021)Click or tap here to enter text. | SiamMask: For object tracking.  Improved Social-LSTM: For trajectory prediction.  PNPoly Algorithm: For predicting unsafe behavior. | The model was tested using video data from the Wuhan Metro project captured through a real-time monitoring system equipped with CCTV cameras. | Prediction involved testing if predicted trajectories entered hazardous areas using the PNPoly algorithm. | mIOU, mAP @ 0.5 IOU, and mAP @ 0.7 IOU were used to evaluate tracking precision |
| Zhu, F., Shao, L., Xie, J., & Fang, Y. (2016) | Handcrafted Representations:  Histogram of Gradients (HOG), Histogram of Optical Flow (HOF)achieved high accuracy.  Learning-Based Representations:  CNN, 3D CNNs, Two-stream CNNs, and Hybrid models. | UCF-101,  HMDB-51,KTH,  Sports-1M. | Utilize spatial and temporal data for more robust and generalized feature extraction.  Depend on large datasets and computational resources. | IDT + FV: Achieved 85.9% accuracy on UCF-101 and 57.2% on HMDB-51.  IDT + HSV: Improved accuracy to 87.9% on UCF-101 and 61.1% on HMDB-51.  Learning-Based Approaches:  Two-Stream CNN: Scored 88.0% on UCF-101 and 63.2% on HMDB-51.  TDD + FV: Combined with Improved Dense Trajectories, it achieved 91.5% on UCF-101 and 65.9% on HMDB-51. |
| Xu, S., Sun, M., Fang, W., Chen, K., Luo, H., & Zou, P. X. W. (2023) | Bayesian-based Knowledge Tracing (BKT) model | Over 1000 on-site photos and videos.  A question bank containing 139 questions related to 35 knowledge concepts tailored to scaffolders. | Includes demographic information, job trades, learning motivations, and cognitive styles. | Accuracy of 80.9 |
| Shao, B., Hu, Z., Liu, Q., Chen, S., & He, W. (2019) | Frequency analysis, correlation coefficient analysis, and variance analysis to study fatal accident patterns in building construction | 2,348 fatal accidents reported between 2012 and 2016. | Falls account for more than 55% of fatal accidents, followed by "struck by object," "collapse," and "hoisting damage." | Fatal Accident Frequency: Distribution of accidents across time, region, and type.  Mortality Rate per GDP: Indicator to assess economic and safety implications. |
| Li, H., Wu, D., Zhang, W., & Xiao, C. (2024) | YOLO-PL: An improved, lightweight variant of YOLOv4 | SHWD (Safety Helmet Wearing Detection)  SHD: Safety helmet detection dataset with 5000 images.  MHD: Motorcycle helmet detection dataset with 3052 images. | Optimization for detecting small safety helmets.  Improved robustness against occlusions and environmental noise. | Results on SHWD Dataset:  AP50: 94.23%.  AP75: 59.00%.  Recall: 94.26%. |
| Wei, R., Love, P. E. D., Fang, W., Luo, H., & Xu, S. (2019) | Spatial and Temporal Attention Pooling Network (ASTPN) | A custom-created video dataset from construction sites, featuring 12 pairs of surveillance videos.  Pre-trained on the publicly available iLIDS-VID database containing 600 sequences of 300 individuals. | Removes redundant video data and focuses on relevant features.  Uses Siamese network and Euclidean distance to recognize individuals. | Achieved an average accuracy of 79.2% using k-fold cross-validation.  Fine-tuned model accuracy: 75%, 75%, and 87.5% across different tests.  Non-fine-tuned model accuracy: Averaged 50%. |
| (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% |