|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Reference | Dataset | Method | Result | Limitation |
| Mukherjee et al. | Webcam videos capturing.  mid-air finger writing | Faster R-CNN for hand detection  Geom. analysisfor finger counting  DWCE for finger-tip detection and tracking  Velocity-based criterion for termination | Precision: 73.1% Real-time: 18.5 fps Char. recognition: 96.11% Incep-v4: 98.12% Incep-v3: 97.96% ResNet: 97.29% VGG-16: 96.75% AlexNet: 96.11% | Challenges in initialization and termination  Precision affected by small fingertip size |
| Alam et al. | Self-collected  RTD datase | LSTM(with normalization) | 99.17% | Misclassification of similar-looking characters, accuracy impacted by RGB webcam limitations , and reduced performance when fingertip faces the camera. |
| LSTM(without normalization) | 98.68% -99.08% |
| CNN (with normalization) | 99.06% |
| CNN (without normalization) | 98.26% - 98.89% |
| 6D motion gesture (6DMG) dataset | LSTM(with normalization) | 99.32% |
| LSTM(without normalization) | 98.68% - 99.08% |
| CNN(with normalization) | 99.06% |
| CNN(without normalization) | 98.26% - 98.89% |
| Chayti Saha et al. | BanglaLekha-Isolated | CNN model with preprocessing steps (resizing, gray scaling, augmentation) | Training : 99.36\%,  Validation : 97\%,  Test : 97.30\% | Dataset size, variability in hand gestures, and limited computational resources may affect generalization and real-time performance |
| Mahmud et al. | Uppercase English letters, 640x480 | Depth-based SVM with DTW distance | 96.85% accuracy | Small dataset (15 users), limited to depth-based features |