Hand Gesture Recognition in Video with Single Short Detector

MADDURI HARIKA
Post Graduate Student, M.C.A Department of Information Technology, Jawaharlal Nehru Technological University, Hyderabad
madduriharika3695@gmail.com

ABSTRACT

With the continuous growth of the global population, innovative human-computer interaction (HCI) technologies are becoming increasingly important in improving quality of life. Among these, gesture-based systems stand out for their ability to enhance accessibility, safety, and user experience, particularly for individuals with physical impairments, while also benefiting the wider community. However, recognizing gestures from video data remains a complex challenge due to variations in motion patterns across different users.This research makes use of the Hand Gesture Classification dataset to examine and compare multiple algorithms for gesture recognition. For the classification task, several approaches were implemented, including Convolutional Neural Networks (CNN) integrated with Support Vector Machines (SVM), Deep Belief Networks (DBN) with SVM, Histogram of Oriented Gradients (HOG) with SVM, Histogram of Optical Flow (HOF) with SVM, as well as an ensemble strategy that combines Xception, CNN, SVM, and a Voting Classifier (utilizing Boosted Decision Trees and Random Forest). For gesture detection, the YOLO family of models was applied, specifically YOLOv5x6, YOLOv5s6, YOLOv8n, and YOLOv9n.The experimental findings reveal that the Xception-based CNN ensemble delivers the highest accuracy, outperforming other models and proving to be particularly effective in reliable gesture recognition from video sequences.

KEYWORDS

Hand Gesture Classification,Video Data,Convolutional Neural Networks (CNN), Support Vector Machines (SVM),Deep Belief Networks (DBN), Voting Classifier

INTRODUCTION

Gesture recognition has become one of the fastest-growing areas in human–computer interaction (HCI). Instead of relying only on traditional input devices like a keyboard and mouse, gesture-based systems allow people to communicate with machines in a more natural and intuitive way. Hand gesture recognition, in particular, plays a major role in this field since it interprets user actions based on the movements of their hands.This technology is now widely applied in several domains such as healthcare, surveillance, and interactive systems, where accurate recognition of gestures is important. Alongside speech recognition, hand gestures are increasingly being studied as natural communication methods. For instance, some researchers have experimented with methods that first remove unwanted background from video frames, convert images to HSV color space, and then apply processes like dilation, erosion, filtering, and thresholding. The processed frames are later classified using algorithms such as Support Vector Machines (SVMs) to identify the performed gestures.Gesture recognition not only improves convenience for all users but also provides accessibility for people with disabilities, making technology more inclusive. The main difficulty, however, lies in the wide variation of motion patterns among different individuals, which makes designing a reliable and accurate recognition system a challenging task.

LITERATURE REVIEW

1. Vision-Based HGR Review: Reviewed 108 studies (2012–2022); identified challenges in segmentation, tracking, and classification; accuracy ranged from 68% to 97%.
2. YOLOv3 Model: Achieved high real-time accuracy (97.68%) without preprocessing; effective in low-resolution and complex environments.
3. Kinect for Elderly Care: Used Kinect V2 and CNNs to detect five essential gestures for elderly assistance; alerts sent via GSM.
4. 2D/3D Gesture Recognition: Used WT, EMD, ANN, and CNN; CNN achieved 100% accuracy in 2D/3D gesture classification.
5. Sign Language Recognition: Proposed a deep learning system combining local/global features and sequence modeling; outperformed state-of-the-art methods.
6. Key Challenges: Hand segmentation, gesture variability, non-rigid motion, and performance in uncontrolled settings remain major issues.

METHODOLOGY

Dataset Description :

Dataset Source: SKIG dataset

Preprocessing steps:

- Data Collection
- Imageprocessing
- Training and Testing
- Modelling
- Predicting

Model Architecture

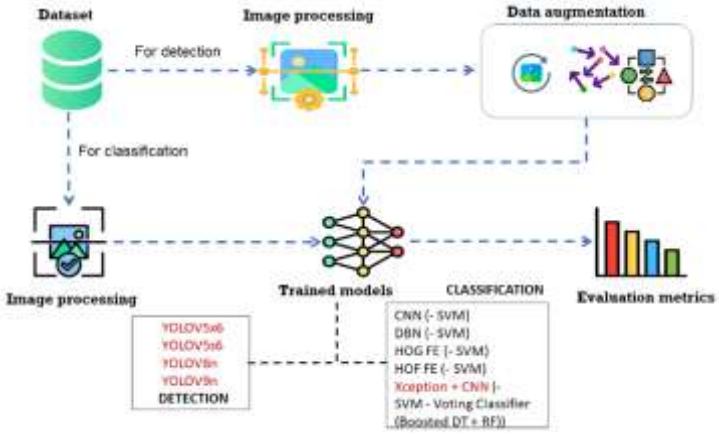


Fig. 1: System Architecture

The system consists of the following modules:
Data Sources – Captures video data from cameras and gesture datasets.
Preprocessing Module – Performs image augmentation, resizing, masking, and feature extraction..

Machine Learning Engine – Uses CNN, DBN, HOG, HOF, Xception ensemble for classification and YOLO models for detection.

Analytics Engine – Detects, classifies, and evaluates gestures in real time.

Dashboard Interface – Flask-based platform for user login, uploads, and viewing results.

Feedback Loop – Improves accuracy by learning from errors and updating models.

System Setup :

Operating System: Windows 10 or above was used as the development and testing environment.

Processor: An Intel Core i5 or higher processor ensured smooth execution of the models.

RAM: At least 8 GB of memory was required for handling data processing and training tasks.

Software & Libraries: Python, Flask, Jupyter, TensorFlow, scikit-learn, SQLite

Frontend: HTML5, CSS3, JavaScript, Bootstrap 4, Jinja2

Development Tools: VS Code, PyCharm, Anaconda

Machine Learning Models Used: CNN, DBN, HOG, HOF, Xception ensemble, YOLO (v5, v8, v9)

Additional Tools: Pandas, NumPy, Matplotlib, and Plotly.

RESULTS

PERFORMANCE EVALUTION FOR CLASSIFICATION

	ML Model	Accuracy	Precision	Recall	F1_score
0	HOG-SVM	0.226	0.193	0.226	0.190
1	HOF-SVM	0.064	0.063	0.064	0.050
2	DBN-SVM	0.111	0.103	0.111	0.092
3	CNN	0.821	0.822	0.821	0.820
4	Xception + CNN	0.956	0.956	0.956	0.956
5	CNN-SVM	0.801	0.801	0.801	0.801
6	Ensemble-SVM	0.945	0.945	0.945	0.945
7	Ensemble-Voting	0.961	0.962	0.961	0.961

Table-1

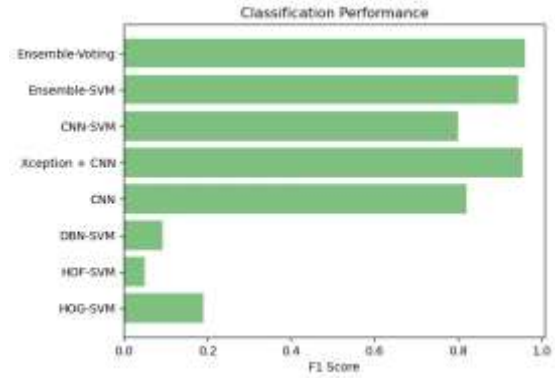
PERFORMANCE EVALUTION FOR DETECTION

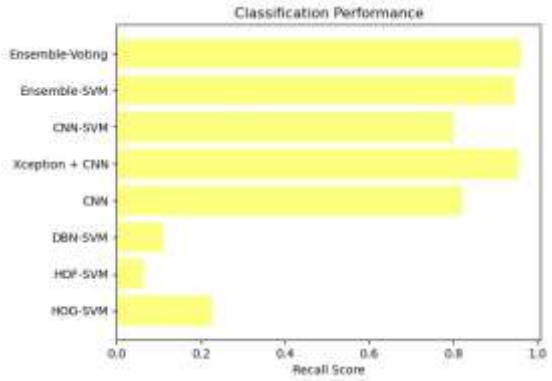
	ML Model	Precision	Recall	mAP
0	Yolo v5s6	0.987	0.990	0.991
1	Yolo v5x6	0.985	0.991	0.992
2	Yolo v8	0.985	0.987	0.991
3	Yolo v9	0.984	0.987	0.993

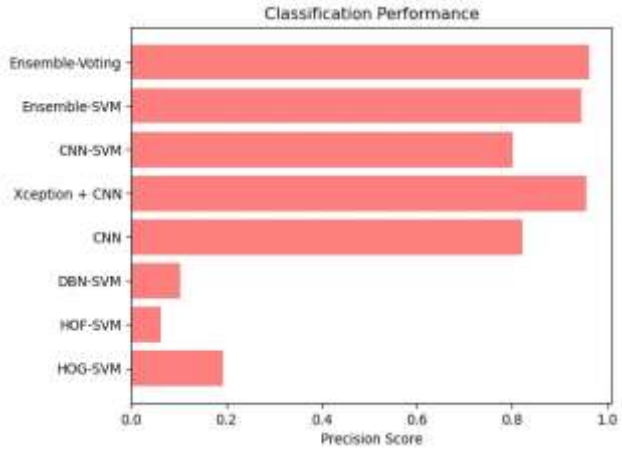
Table -2

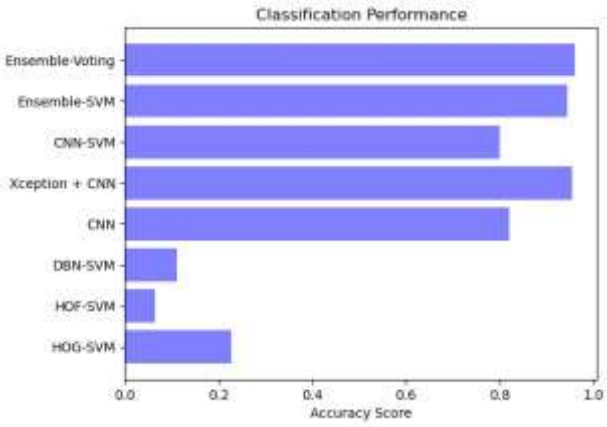
Graphs:

COMPARISION GRAPHS FOR CLASSIFICATION









COMPARISION GRAPHS FOR DETECTION:

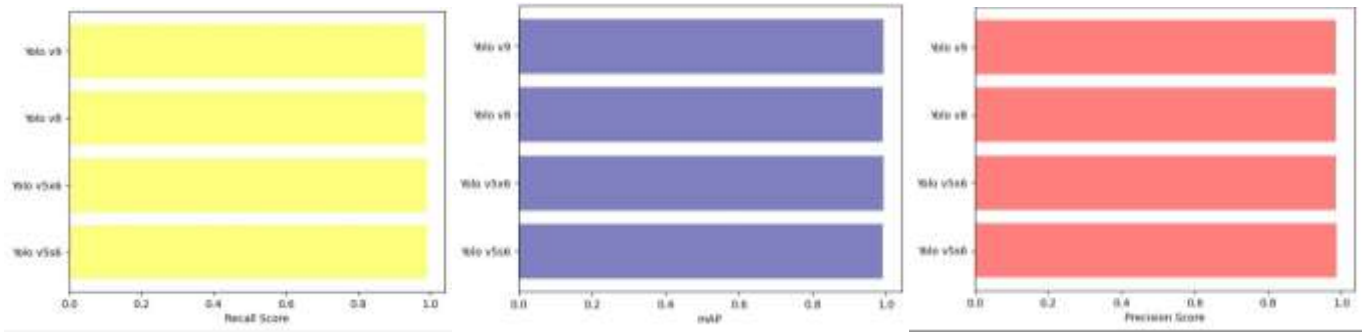


Fig. 2:Comparision Graphs for Classification and Detection

OUTPUT SCREENS

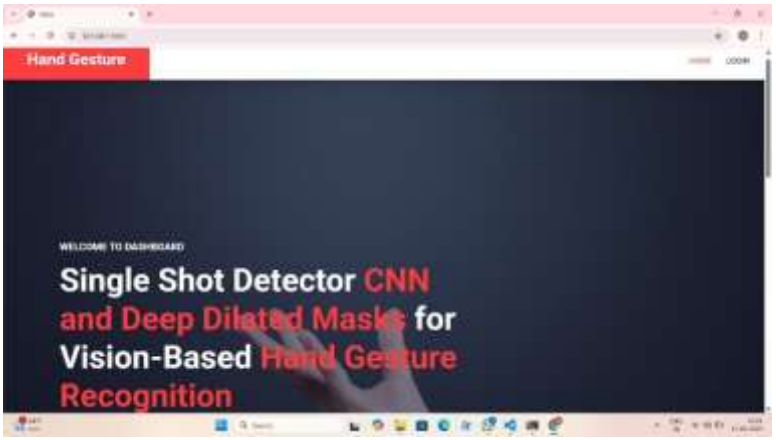


Fig. 3: Home Screen Interface

Fig. 4: User Registration Form

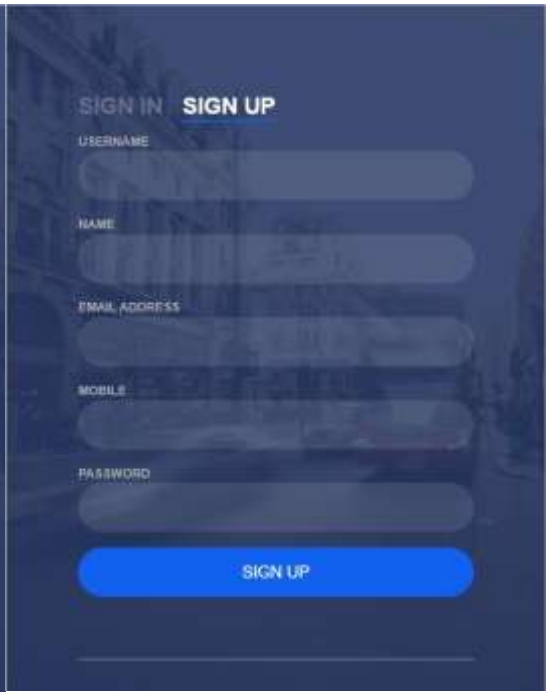


Fig. 5: Login Page

Classification:

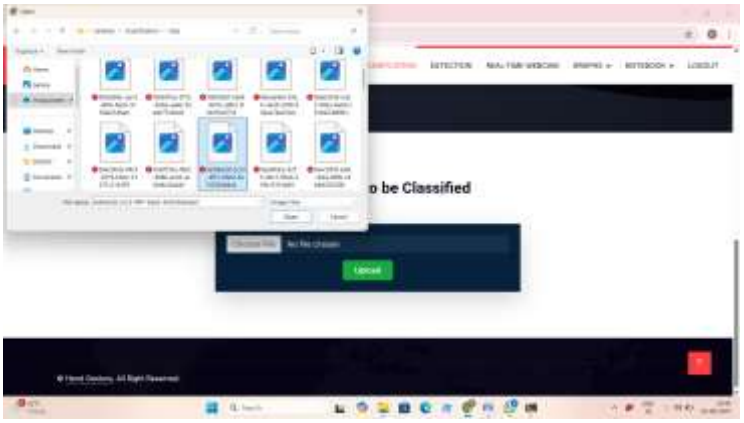


Fig. 6: Input Page for classification

Fig.7: Results Page for classification

Detection:

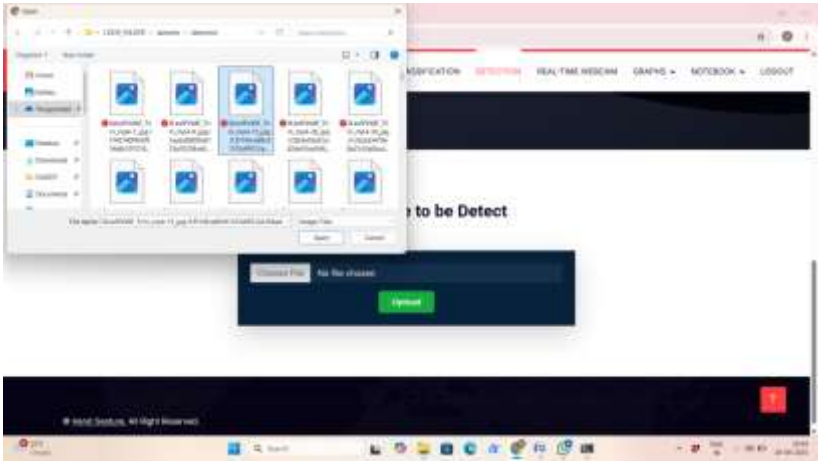


Fig. 8:Input Page for detection

Fig.9: Results Page for detection

CONCLUSION

The proposed hand gesture recognition system marks a major step forward in the field of human-computer interaction. By utilizing the Xception model in combination with Convolutional Neural Networks (CNNs), the system delivers high accuracy in identifying gestures from video sequences. This approach effectively tackles challenges related to individual variations in movement, making the recognition process more consistent and dependable. Moreover, the inclusion of advanced detection methods from the YOLO framework enables real-time gesture tracking and localization, further improving performance. Beyond technical achievements, this system holds great promise for assistive technology applications, particularly for individuals with disabilities, while also enhancing everyday user interaction with digital devices. The integration of CNN and Xception demonstrates strong potential for improving accessibility, usability, and overall user engagement. Thus, the system represents meaningful progress toward developing intuitive, efficient, and user-friendly gesture-based interfaces.

REFERENCES

1. F. A. Farid, N. Hashim, J. Abdullah, M. R. Bhuiyan, W. N. S. M. Isa, J. Uddin, M. A. Haque, and M. N. Husen, ‘‘A structured and methodological review on vision-based hand gesture recognition system,’’ J. Imag., vol. 8, no. 6, p. 153, May 2022.
2. A. Mujahid, M. J. Awan, A. Yasin, M. A. Mohammed, R. Damaševišius, R. Maskeličnas, and K. H. Abdulkareem, ‘‘Real-time hand gesture recognition based on deep learning YOLOV3 model,’’ Appl. Sci., vol. 11, no. 9, p. 4164, May 2021.
3. M. Oudah, A. Al-Naji, and J. Chahl, ‘‘Elderly care based on hand gestures using kinect sensor,’’ Computers, vol. 10, no. 5, 2021. [Online]. Available: https://doi.org/10.3390/computers10010005
4. N. Alnaim, ‘‘Hand gesture recognition using deep learning neural networks,’’ Ph.D thesis, Brunel University London, 2020.
5. M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, T. S. Alrayes, H. Mathkour, and M. A. Mekhtiche, ‘‘Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation,’’ IEEE Access, vol. 8, pp. 192527–192542, 2020.
6. Y. Zhang, L. Shi, Y. Wu, K. Cheng, J. Cheng, and H. Lu, ‘‘Gesture recognition based on deep deformable 3D convolutional neural networks,’’ Pattern Recognit., vol. 107, Nov. 2020, Art. no. 107416.
7. M. R. Bhuiyan, D. J. Abdullah, D. N. Hashim, F. A. Farid, D. J. Uddin, N. Abdullah, and D. M. A. Samsudin, ‘‘Crowd density estimation using deep learning for Hajj pilgrimage video analytics,’’ FResearch, vol. 10, p. 1190, Nov. 2021