Hand Gesture Recognition System for ASL Alphabet

SFU SCHOOL OF COMPUTING SCIENCE

Aarushi Bhatnagar, Riya Ali, Bhavya Dintakurthi, Mariam Elamrawy

Project Overview

Our Hand Gesture Recognition System for the ASL Alphabet is built on a fully custom dataset of over 5,000 labeled hand-gesture images, with a complete preprocessing pipeline and a purpose-built Convolutional Neural Network (CNN).

Data Sources

- Fully **self-collected dataset**, tailored for ASL letters.
- Webcam frames serve as both **training data** and **real-time inference input**.
- Validation split ensures the model generalizes across lighting, angles, and hand variation.

Modeling Approach

We utilized Python, integrated OpenCV for image capture, MediaPipe for hand detection, and TensorFlow/Keras for model training.

Outputs & Visualization

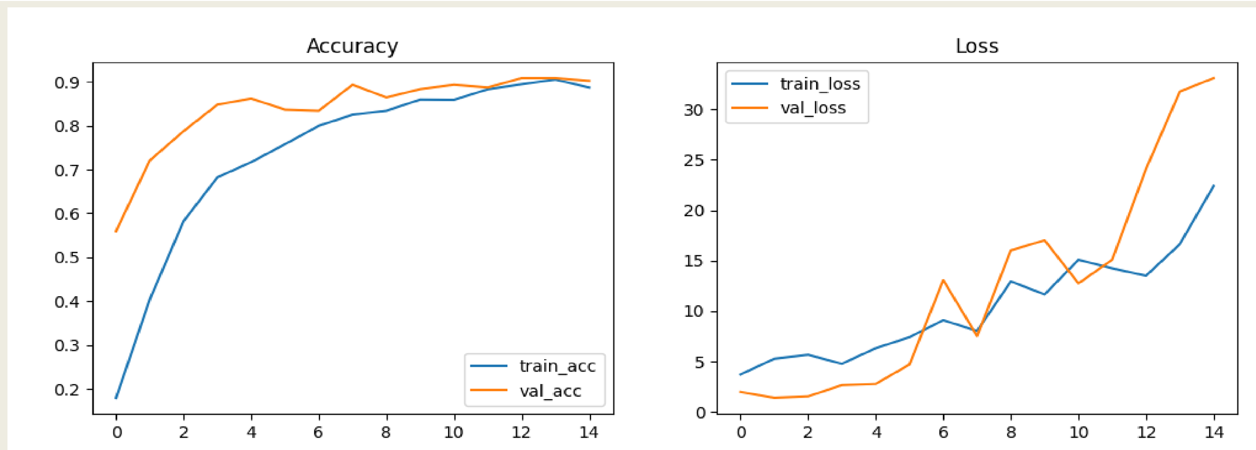
Our model provides real-time ASL letter predictions directly on the webcam feed, displaying both the recognized letter and its confidence score. The interface also shows the cropped hand region and the standardized input image. Overall performance is plotted in training accuracy/loss curves and a confusion matrix,

Motivation & Purpose

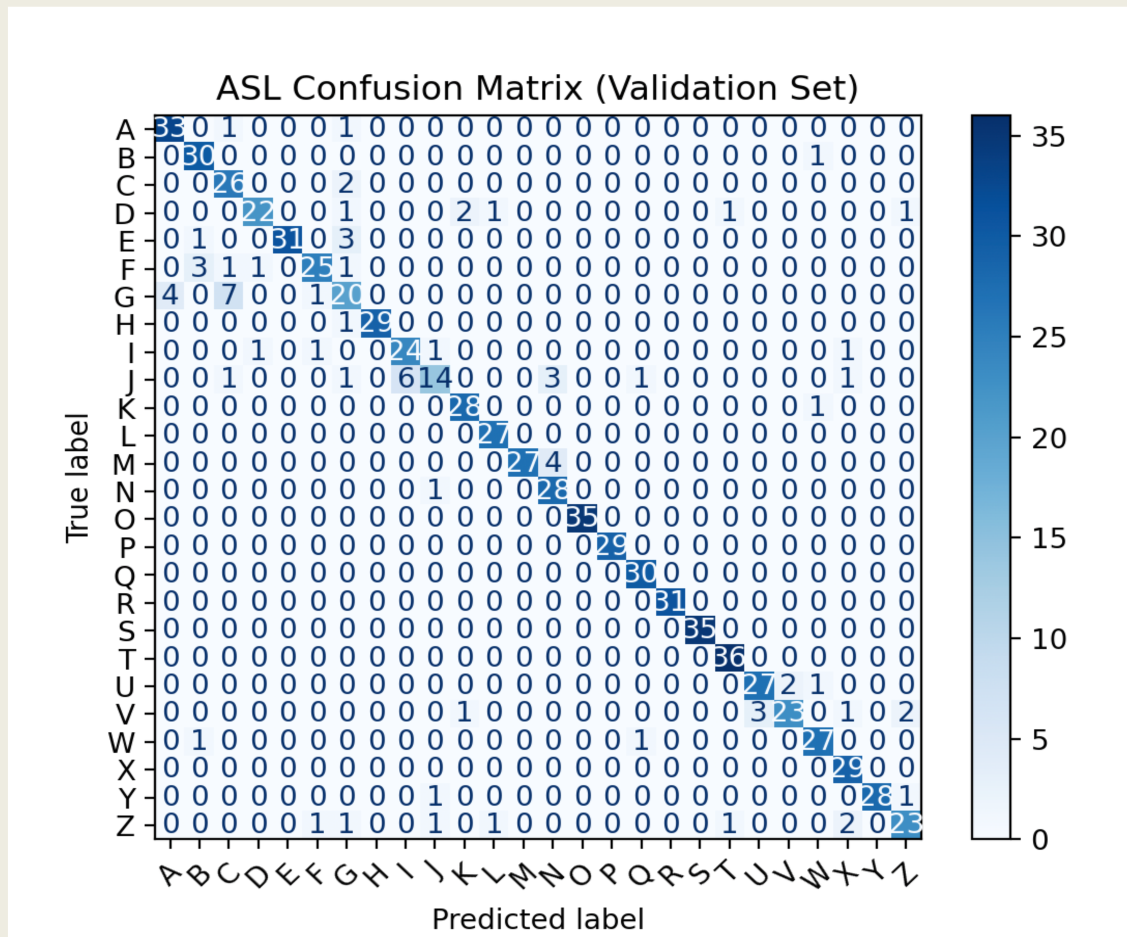
Current ASL recognition systems still face major limitations: real-time gesture detection is computationally demanding, accuracy drops under changing lighting, backgrounds, or poor hand visibility and models struggle to generalize across different users' hand shapes, skin tones, and signing styles. Progress is further restricted by the scarcity of high-quality public ASL datasets as many corpora are corrupted, inaccessible, or heavily licensed due to the difficulty of anonymizing facial and body features. Additionally, over 85%of existing datasets use only a single front-facing camera view, even though real-world signing often occurs from varied angles, reducing model robustness.

Our project addresses these issues by integrating OpenCV and MediaPipe for stable real-time hand tracking and training a CNN on a custom dataset specifically designed to avoid these limitations.

Results



The plot shows a **90% accuracy** by epoch 14, indicating good generalization without overfitting. Loss decreases for both sets early on, with validation loss rising slightly near the end but overall performance remains stable and reliable.

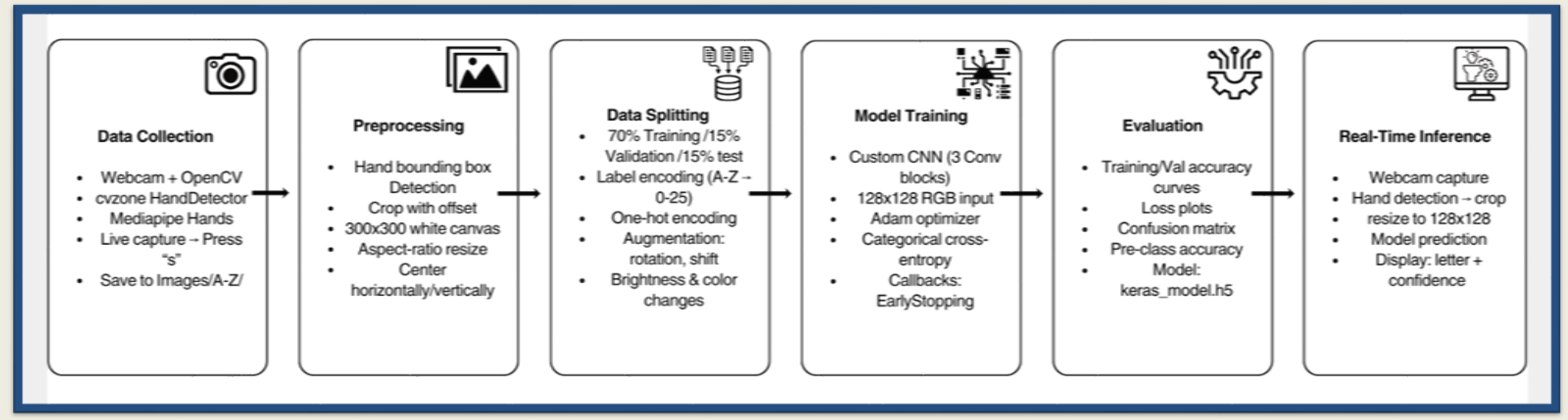


Every class has 25–36 correct predictions out of 30.
Strong Performing Letters: B, H, K, L, O, P, Q, R, S, T, W, X

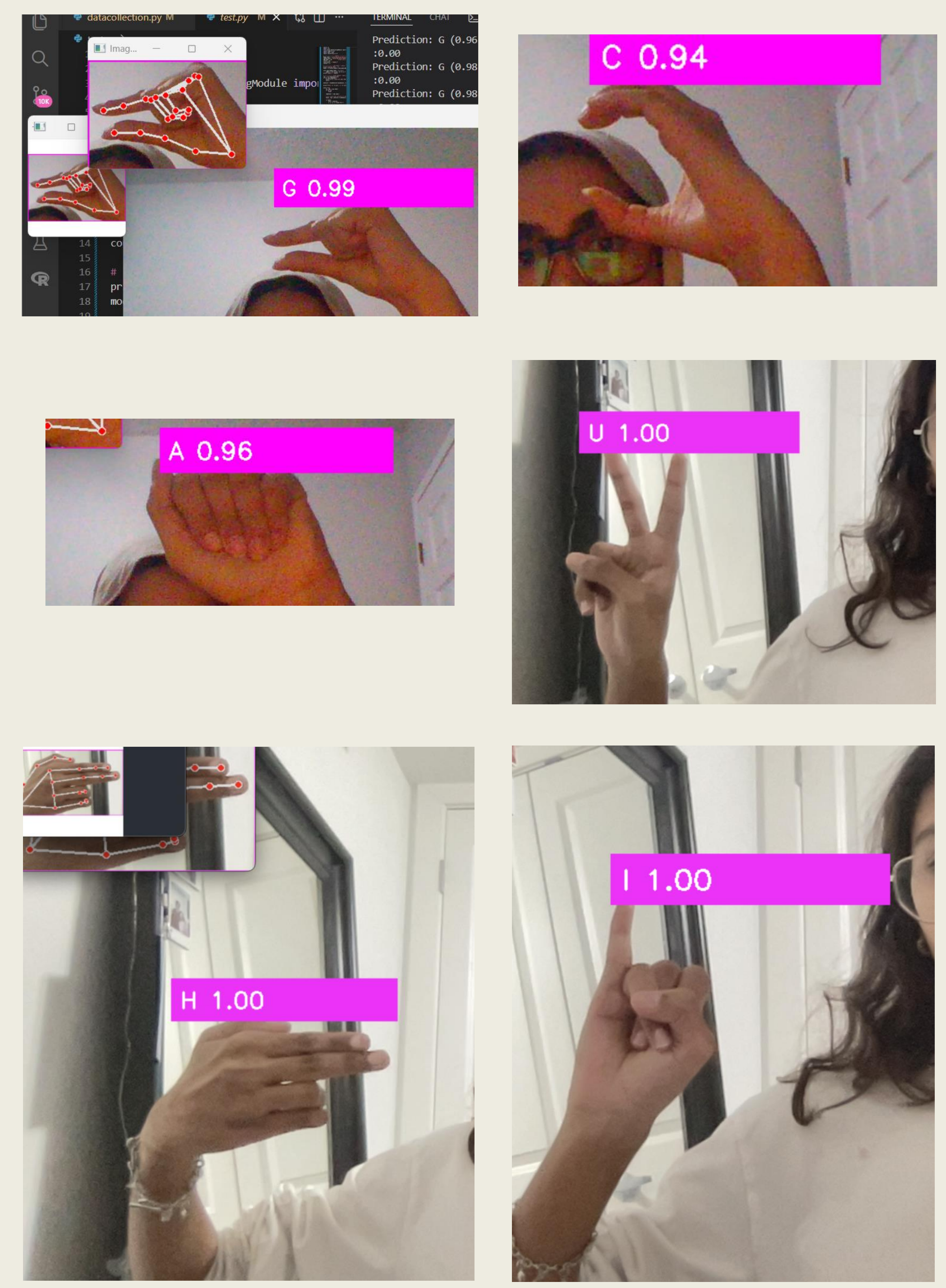
Some confusion exists between visually similar gestures:

- M is classified as A and N
- D is sometimes predicted as X and Z.
- V is sometimes misclassified as U

Data Pipeline



Visualization



Methods & Technologies

Data Collection

- Custom ASL dataset using webcam + MediaPipe
- 200+ images per letter (A–Z)
- Bounding box extraction with aspect-ratio resize
- Centered on 300×300 white background

Model & Training

- Convolutional Neural Network
- Aggressive data augmentation (rotation, shifts, zoom, brightness)
- Confusion matrix generated using scikit-learn

Tools & Libraries

Python, OpenCV, MediaPipe Hands, TensorFlow/Keras, NumPy/Matplotlib, Scikit-learn

References

ML6. (2022, November 6). Challenges for Sign Language Translation. *ML6*. https://www.ml6.eu/en/blog/challenges-for-sign-language-translation

Ohri, A. (2025, April 23). *Real-time American sign language recognition using webcam, computer vision, and machine learning*. California State University, Sacramento. https://scholars.csus.edu/esploro/outputs/graduate/Real-time-American-sign-language-recognition-using/99258206818001671

Arikeri, P. (n.d.). *American Sign Language (ASL) Dataset*. Kaggle.com. Retrieved November 23, 2025, from https://www.kaggle.com/datasets/prathumarikeri/american-sign-language-09az

Bali, M. (2022, April 23). *Sign Language Detection for Deaf using Deep Learning, MediaPipe and OpenCV*. Medium. https://medium.com/@mayank.bali/sign-language-detection-for-deaf-using-deep-learning-mediapipe-u-opencv-4c5151e2374c