Real-Time ASL Gesture Recognition: A Comparative Study

Group Number：B109　　Group Member: Hengrui Li Yusen Fu Chenglong liu

1. INTRODUCTION

Objective: Develop a real-time computer vision system to translate static ASL alphabets (A-J) into text.

Approach: We evaluate **MediaPipe** landmark extraction using both Supervised (Classification) and Unsupervised (Clustering) learning to determine the most effective method.
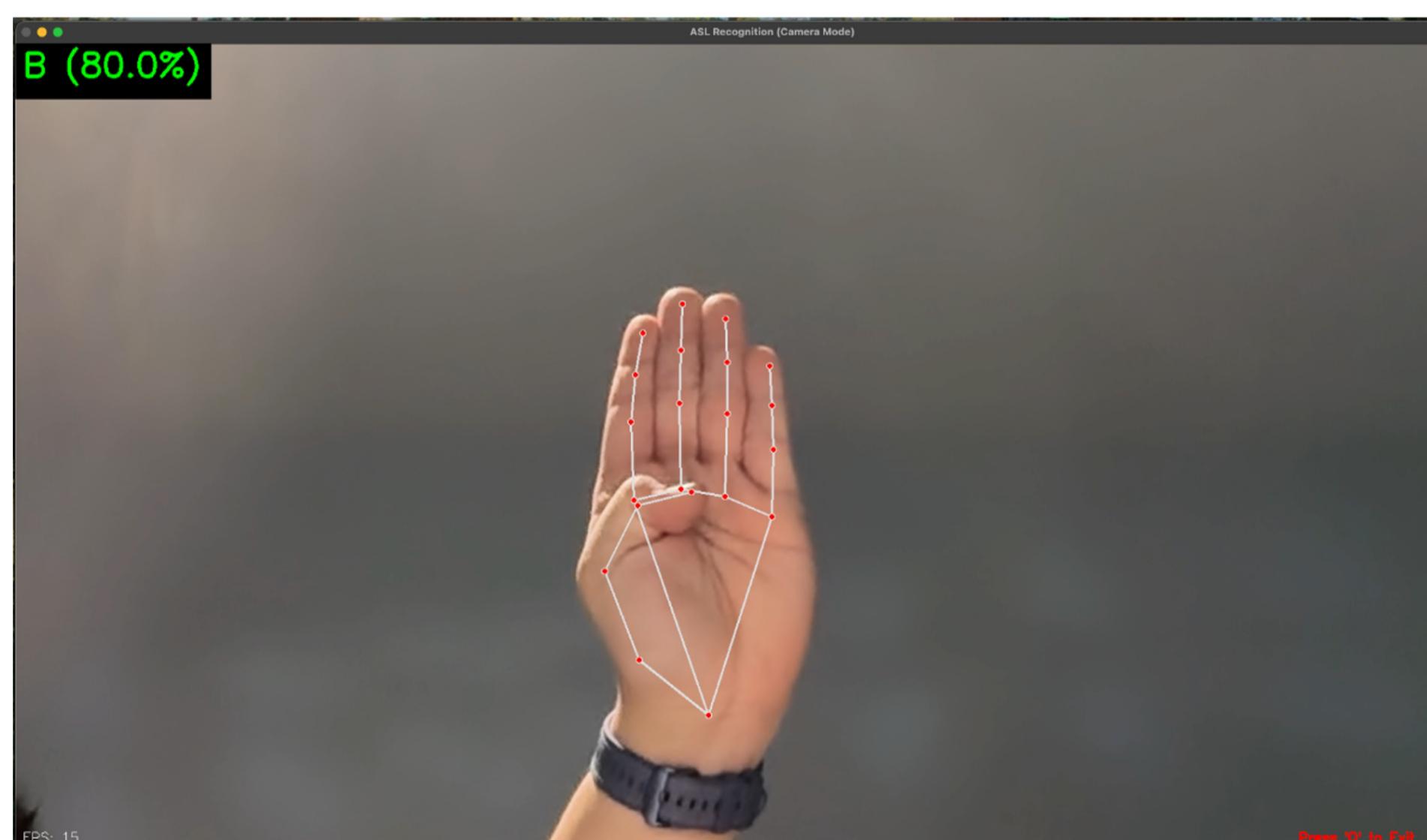
2. METHODOLOGY

1. Feature Extraction
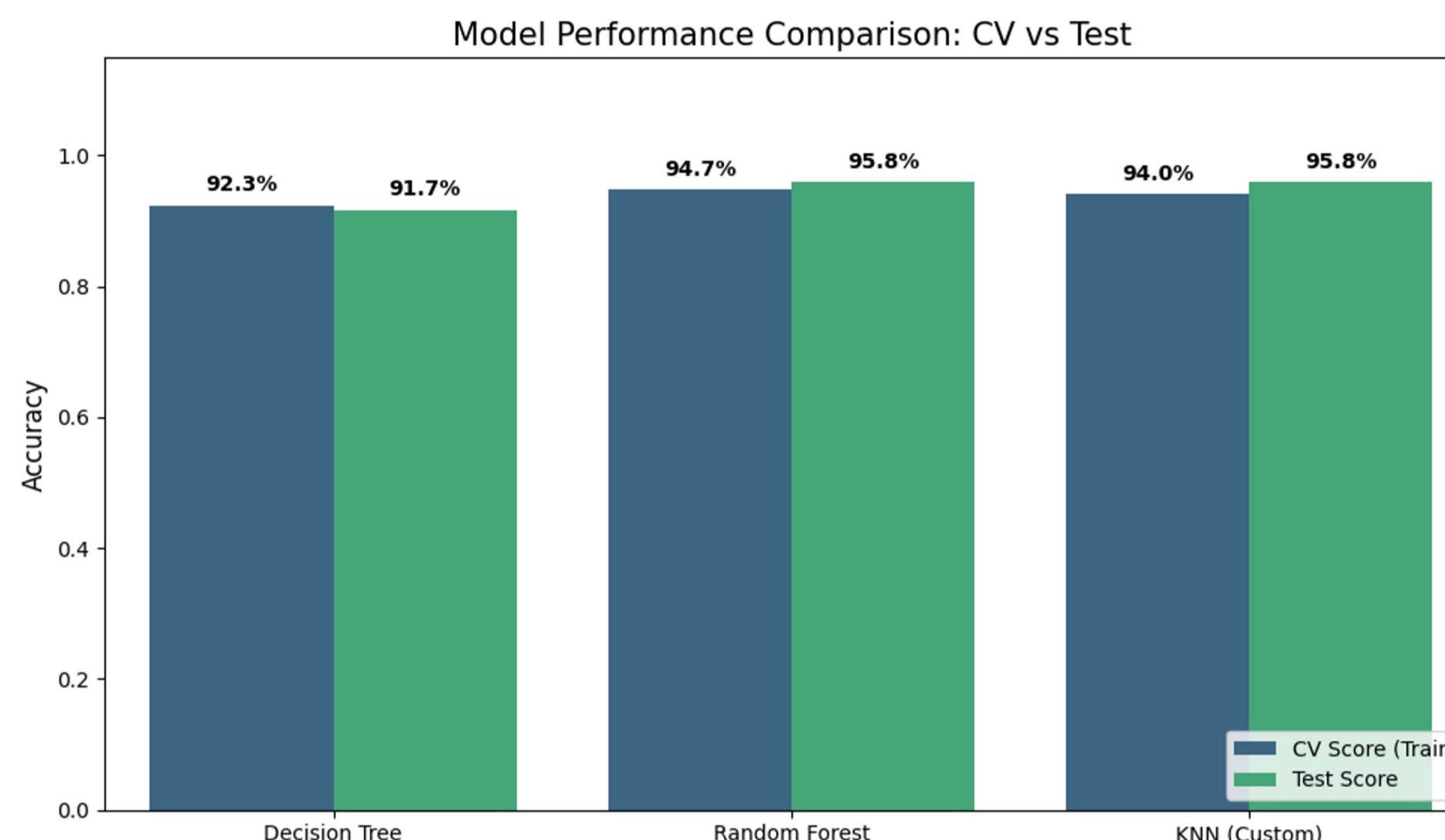
- **Input:** Webcam RGB frames.
- **MediaPipe:** Extracts 21 3D landmarks (x,y,z).
- **Pre-processing:** Wrist-relative normalization [-1, 1] for scale invariance.

2. Algorithms
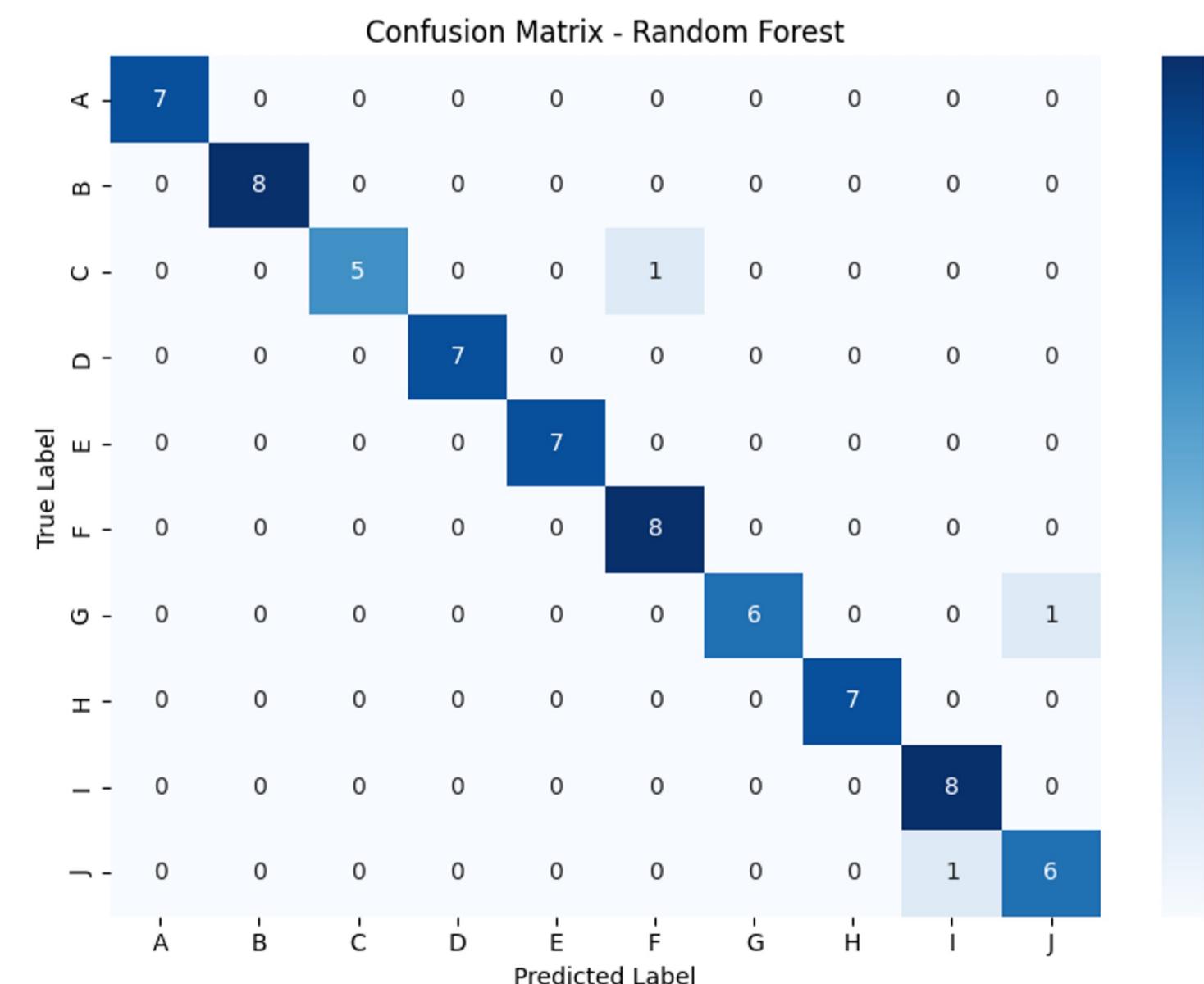
- **Supervised:** Random Forest, Decision Tree, Custom KNN.
- **Unsupervised:** K-Means Clustering (K=10).
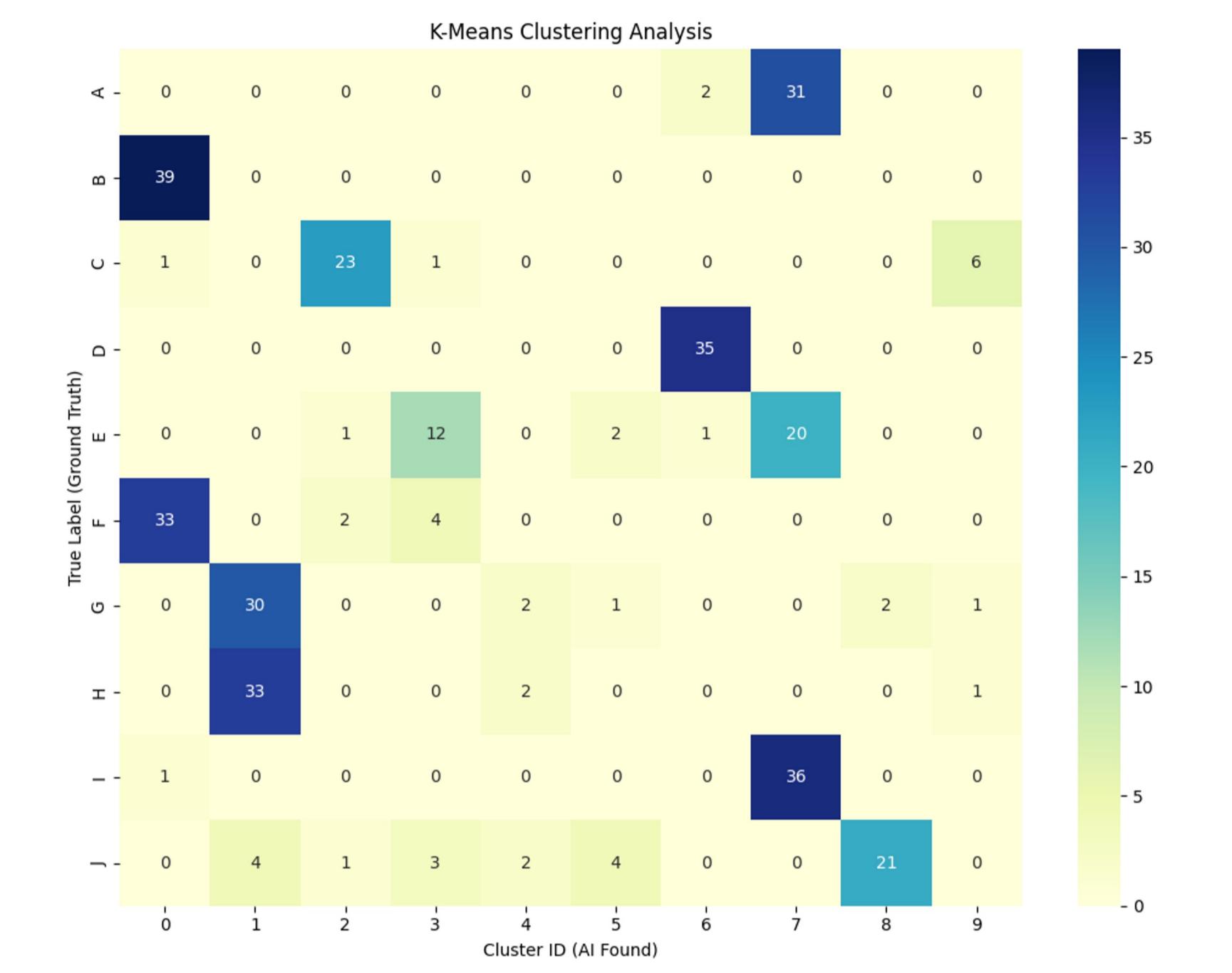
3. RESULTS: MODEL PERFORMANCE

Model Performance Comparison: CV vs Test

- **Random Forest & Custom KNN:** Highest accuracy (**95.8%**).
- **Decision Tree:** 91.7% accuracy.
- **Stability:** Minimal gap between CV and Test scores suggests no overfitting.

Confusion Matrix - Random Forest

4. UNSUPERVISED ANALYSIS

K-Means Clustering Analysis

- **Clustering:** K-Means separated data into clusters matching Ground Truth labels.
- **Insight:** Geometric feature extraction creates a naturally separable feature space without labels.

5. CONCLUSIONS

- **Features:** Geometric landmarks ensure high accuracy.
- **Best Model:** Random Forest (High accuracy & stability).
- **Real-Time:** Efficient webcam performance.

Future: Support dynamic gestures via LSTM.