DEPI

SIGN - 2 - TEXT

Team Number : 232

DEPI

TEAM MEMBERS

01 *Zenab Osama*

02 *Abdelrahman Darwish*

03 *Hana Ibrahim*

04 *Amr Fouad*

05 *Ahmed Saeed*

06 *Mahmoud Hallul*

DEPI

PROJECT OVERVIEW

01

Introduction

start by establishing the importance of Sign Language (SL) and the goal of bridging the communication gap using technology. Define the project's scope: real-time recognition of the manual alphabet (fingerspelling) using Computer Vision.

02

The Initial Challenge

Detail the initial attempt using the widely known SNL (Sign Language) dataset and the CNN (Convolutional Neural Network) architecture. Describe the unexpected failure: High accuracy on static images but poor, random prediction in real-time camera feed. Mention the hypothesis: The variance in the public dataset (lighting, backgrounds, and hand physiology) was too high to generalize effectively to a live camera feed.

03

The Pivotal Shift

Introduce the critical decision to create a custom dataset. Mention the diversity gained from different team members' hands (finger size, length, and shape variations) and how this improved data relevance.

DEPI

PROJECT OVERVIEW

04

Problem Statement

State the final, focused problem: To develop a robust, real-Detail the custom data collection process. Emphasize why this was crucial: The data specifically matches the operational environment (lighting, camera type, angle). Describe the preprocessing steps (e.g., background subtraction/segmentation, normalization, image resizing) necessary to isolate the hand gesture and feed clean data into the model.time computer vision system that accurately translates hand gestures for the alphabet into text, overcoming the generalization limitations of standard deep learning models on live feeds.

05

Data Collection & PreprocThe Successful Model Architectureessing

After spending months experimenting with CNNs, the model suffered from clear overfitting. It performed extremely well on images during testing, but once I switched to the live camera, the predictions became almost random.

This issue was solved when I moved to using MediaPipe for reliable feature extraction and a Random Forest Classifier for stable classification.

This combination delivered much higher efficiency and consistent performance in the live camera feed.

06

Feature Engineering

The model captures key spatial and geometric features of the hand, such as finger orientation and palm position, using precise MediaPipe landmarks.

These structured features allow the classifier to understand the hand's pose more reliably.

This gives it a clear advantage over the previous CNN, which failed to interpret these relationships and predicted randomly in live video.

DEPI

PROJECT OVERVIEW

07

Results and Evaluation

For live detection, the system uses OpenCV for camera input and fast image processing, ensuring a high FPS for smooth and usable real-time translation.

The final model achieved 95%+ accuracy on the live feed, unlike the initial CNN which failed to generalize and produced random predictions.

Key performance metrics include Accuracy, Low Latency (fast predictions), and Robustness across different users and backgrounds.

The main success of the project is the model's ability to reliably generalize to unseen live data — the core goal of real-time sign detection

08

Application and Impact

The project enables text input through hand gestures, making communication more accessible.

It can be extended to full words, commands, and assistive technology for people with speech impairments.

09

Future Work & Conclusion

Next Steps: Expand dataset, add full Sign Language vocabulary, optimize deployment, and track gesture sequences.

Summary: Quality data and the right model matter more than popular architectures or public datasets.

MOTIVATION

Sign Language (SL) is a vital form of communication for millions of individuals globally. Our project is driven by the imperative to leverage modern technology to enhance accessibility, foster inclusion, and provide innovative solutions for the Deaf and Hard of Hearing community. Specifically, we aim to address the need for reliable, accessible, and low-cost methods for real-time translation of the manual alphabet (fingerspelling) into text. By turning any standard computer camera into an effective tool, we seek to enable a natural, hands-free text input method that can significantly improve interaction and communication for users in various settings. This effort aligns with the broader mission of using Computer Vision (CV) to create truly helpful assistive technologies.

THE GENERALIZATION CHALLENGE

Our initial technical investigation involved training Convolutional Neural Networks (CNNs) on large, publicly available datasets, such as the widely known Sign Language dataset (SNL). This approach successfully yielded high accuracy on static test images. However, a critical failure occurred when the model was deployed into the target environment: the live webcam feed. The system exhibited random, unusable predictions in real-time. We identified the core technical problem as a severe lack of generalization. The high variance found in public datasets—stemming from diverse lighting, varied backgrounds, and significant differences in hand physiology (finger size, length, and shape)—did not match our local operating conditions

DEPI

FUTURE TECHNICAL DEVELOPMENT

01

Dataset Expansion and Robustness

Expand the Custom Dataset: Systematically collect data from a significantly larger and more diverse user base (different ages, genders, skin tones) and under varying real-world conditions (different lighting levels and backgrounds) to further boost the model's generalization capabilities.

02

Feature and Scope Enhancement

Transition from Fingerspelling to Full Signs: Develop the model to recognize not just individual letters (fingerspelling) but complete, common Sign Language vocabulary words. This would require integrating temporal analysis (tracking the movement and sequence of signs over time).

03

Model Optimization and Speed

Edge Deployment and Optimization: Re-train and optimize the model architecture (e.g., quantize the model, use lightweight architectures like MobileNet/TinyML) to ensure efficient deployment on lower-power devices such as mobile phones or embedded systems, significantly reducing latency.

04

User Interaction Improvements

Integration with Text Services: Integrate the output with instant messaging applications or live transcription services to make the translation directly usable in daily communication contexts.

FUTURE MARKETING & COMMERCIALIZATION

01

Productization Strategy

Develop a Standalone Mobile Application (MVP): Package the system into an easy-to-use mobile application (iOS/Android) marketed as an "Instant Sign-to-Text Translator," making the technology highly portable and accessible.

02

Target Market Penetration

Partnering with Educational Institutions: Market the tool to schools and universities as an assistive learning device for students with hearing difficulties or for teaching Sign Language.

03

Corporate Social Responsibility (CSR)

Enterprise Integration: Approach corporations and public service providers (banks, hospitals) to integrate the system into their service desks or customer kiosks to ensure accessibility compliance and better serve the deaf community

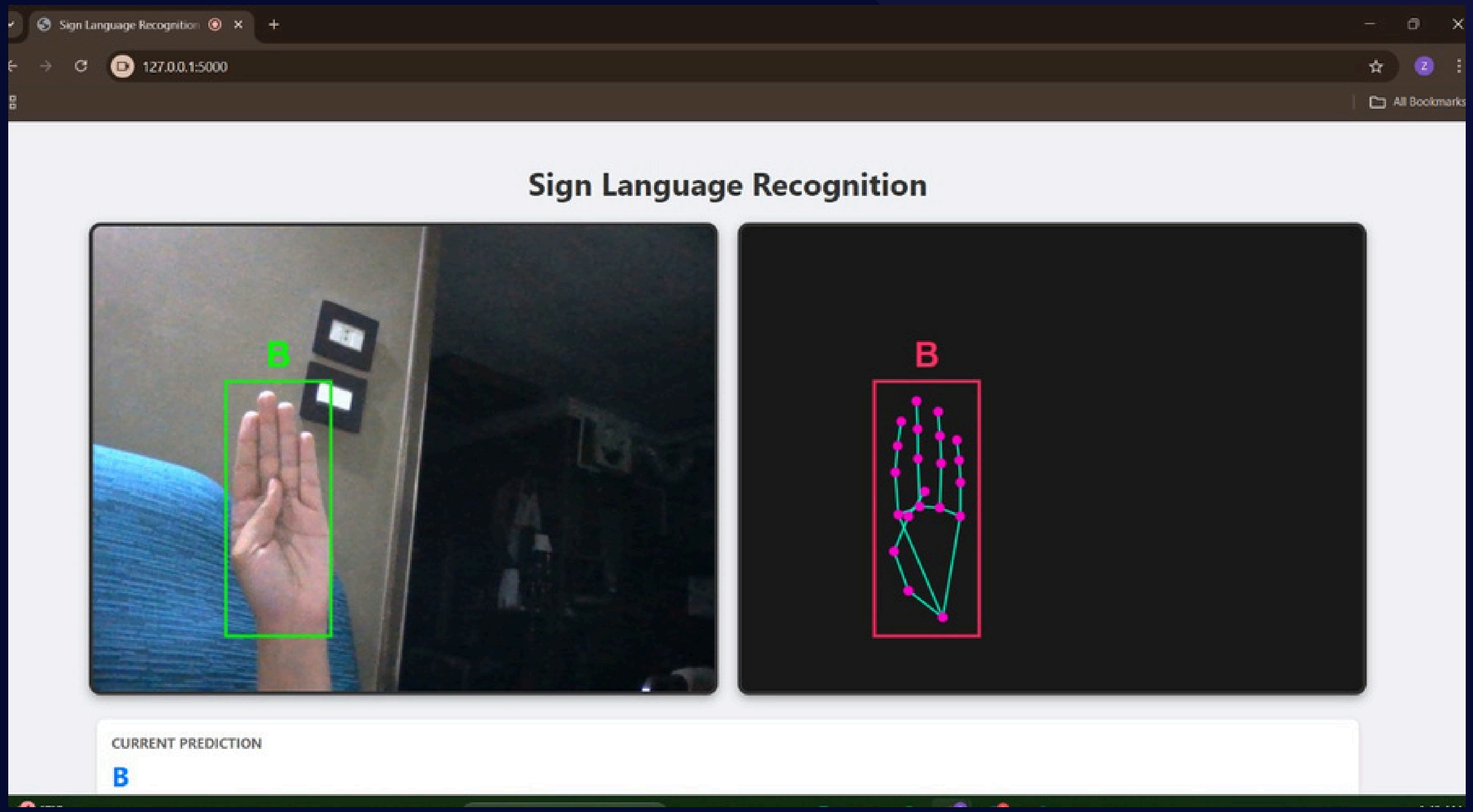
04

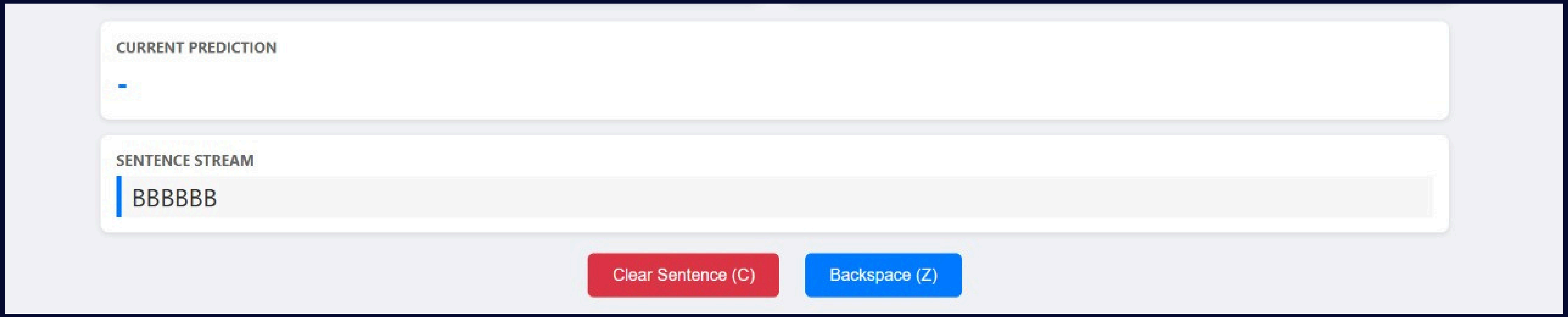
Monetization Model

Freemium Model: Offer the basic alphabet recognition as a free service, with premium features (e.g., full-word sign recognition, personalized hand customization/training, ad-free experience) available through a subscription.

DEPI

OUTPUT


Sign Language Recognition
127.0.0.1:5000
Sign Language Recognition
B
B
CURRENT PREDICTION
B


CURRENT PREDICTION
-
SENTENCE STREAM
BBBBBB
Clear Sentence (C)
Backspace (Z)

Thank You