Sign Language Recognition Using Machine Learning

Hasan Abbas
Reg No: 2022204
u2022204@giki.edu.pk

Muhammad Murtaza
Reg No: 2022402
u2022402@giki.edu.pk

***Abstract*—*This project focuses on developing a Sign Language Recognition system capable of interpreting hand gestures rep- resenting numbers (0-9) and alphabets (A-Z). By leveraging machine learning models and a robust dataset of hand gestures, the system predicts the intended character with high accuracy. The models used for this project include a pre-trained logistic regression model and a scaler for feature normalization. Streamlit is employed for the user interface, enabling a seamless interaction experience, while voice interaction functionality vocalizes the predicted character, further enhancing accessibility.*

Index Terms—Sign Language, Machine Learning, Logistic Regression, Streamlit, Accessibility

I. Paper Organization

The report begins with an **Introduction** outlining the project's objectives and scope. The **Background** section discusses the motivation behind the project and relevant literature. **Methods** describes the dataset, preprocessing steps, and the models employed. **Results** showcase the system's performance and provide observations on its efficiency. **Alternative Methods** explores other approaches considered. Finally, the **Conclusion** summarizes the project findings and discusses future enhancements.

II. Background

A. Motivation

Sign language serves as a critical communication tool for mute individuals. However, the reliance on interpreters or specific knowledge of sign language by others limits accessibility. This project aims to address this gap by developing an automated system capable of recognizing and translating hand gestures into spoken characters or numbers. By bridging this communication gap, the system aspires to enable real-time, barrier-free interaction for mute individuals. While currently a semester project, future work aims to advance this concept towards a comprehensive communication tool.

B. Literature Review

Hand gesture recognition has been explored extensively in recent years due to its potential in various applications. Traditional approaches include Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs). While these methods achieve high accuracy, their computational requirements often pose challenges for real-time use on resource-constrained devices. Logistic regression, with its simplicity and efficiency, offers a compelling alternative. For this project, a supervised learning approach was adopted, utilizing labeled images and extracted hand landmarks as input features.

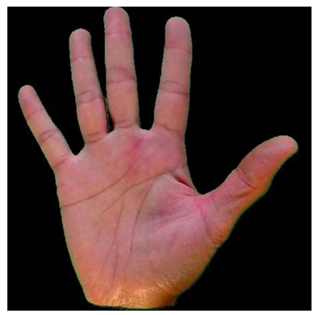

Fig. 1. Sign language for 5

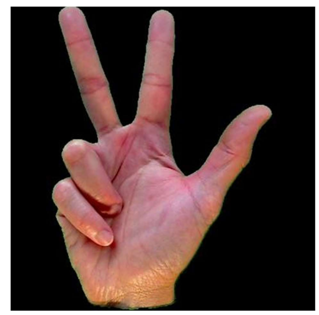

Fig. 2. Sign language for 3

III. METHODS

A. Dataset Description

The dataset used in this project consists of two primary components:

- **Image Folders:** These folders contain labeled images representing hand gestures for numbers (0-9) and alphabets (A-Z). (figures 1 and 2)
- **CSV File:** A CSV file containing 63 extracted hand landmarks serves as the structured input for model training and prediction.

The combination of these resources ensures that the model is trained on both raw images and structured numerical data, enhancing its predictive capabilities.

B. Preprocessing

- **Landmark Extraction:** MediaPipe's hand solution extracts 63 hand landmarks from each image, representing critical points of the hand. Figure 3 shows landmarks in one hand having 21 landmarks, for three dimension makes a total of 63.
- **Normalization:** The landmarks are normalized relative to the dimensions of the limiting box, ensuring consistency between images of varying sizes.

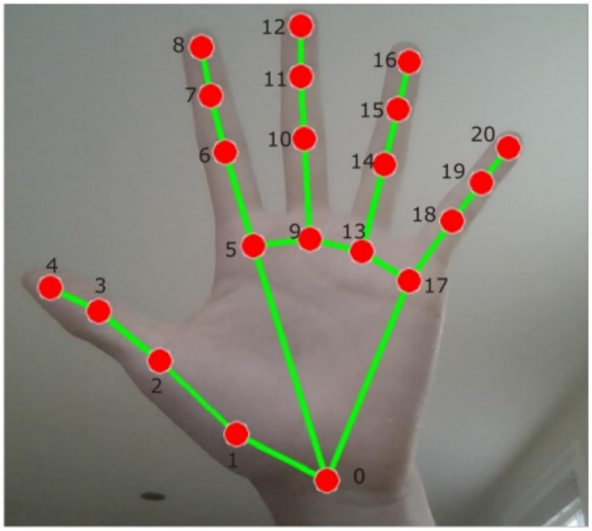

Fig. 3. landmarks of one hand

C. Models

- **Logistic Regression:** This model was selected for its balance of computational efficiency and accuracy, making it ideal for real-time applications.
- **Scaler:** Standardization of input features using a scaler ensures optimal model performance.

D. Implementation

The system is implemented in Python, utilizing the following libraries:

- **MediaPipe:** For hand landmark detection.
- **OpenCV:** For image processing and visualization.
- **Joblib:** To load pre-trained models efficiently.
- **Streamlit:** Provides an intuitive graphical user interface.
- **SpeechRecognition and pyttsx3:** Enables voice output of the predicted character or number.

IV. Results

The system exhibits exceptional hand gesture recognition capabilities, adapting seamlessly to varying lighting conditions and user differences. The logistic regression model operates efficiently, delivering real-time predictions with minimal latency, even when faced with noisy inputs. Figures 4 and 5 showcase the Streamlit-based interface, offering users an interactive platform to engage with and test the system.

Voice Interaction: The predicted character or number is vocalized using pyttsx3, enhancing accessibility for visually impaired users. This feature ensures inclusivity and adds a dynamic dimension to the system's functionality.

V. Alternative Methods

A. Convolutional Neural Networks (CNNs)

CNNs are well-suited for complex image recognition tasks and could provide higher accuracy for gesture recognition. However, due to their computational intensity, they are less suitable for real-time applications on low-resource devices. Additionally, the project's constraints did not allow for an in-depth exploration of CNNs as they were outside the semester syllabus.

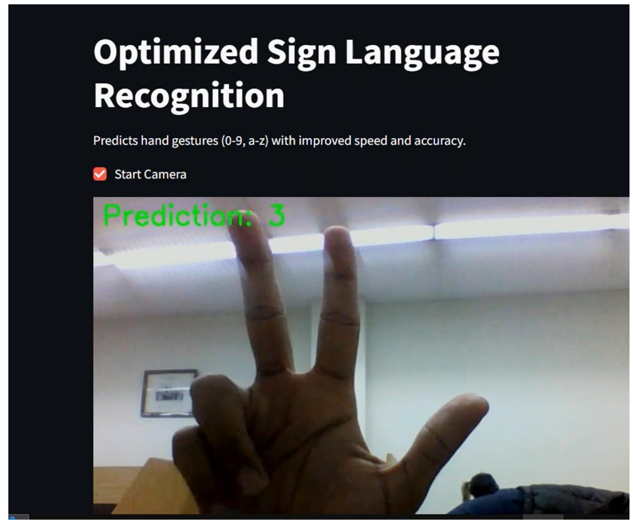

Fig. 4. Predicted number 3

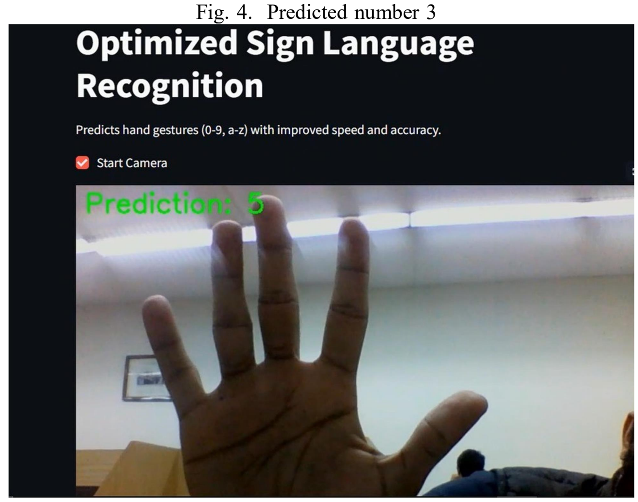

Fig. 5. predicted number 5

B. Random Forests

Random Forest models offer interpretability and can achieve good accuracy. However, they lack the speed and simplicity of logistic regression for this specific task. Their use was explored but ultimately not adopted due to these limitations.

VI. Conclusion

This project successfully developed a Sign Language Recognition system that leverages logistic regression and MediaPipe for accurate and efficient gesture recognition. The addition of voice interaction further enhances accessibility, making the system user-friendly and inclusive. Future work will focus on extending the dataset to include dynamic gestures and exploring deep learning models, such as CNNs, for more complex tasks. This effort underscores the potential of AI-driven solutions in bridging communication gaps for the mute community.

VII. Acknowledgements

We thank our instructor, Dr. Khurram Jadoon, for his guidance and support throughout this project. Additionally, we acknowledge the contributions of open-source libraries such as MediaPipe and Streamlit, which played a pivotal role in this work.

VIII. References

1) MediaPipe Documentation: https://mediapipe.dev/
2) Scikit-learn Documentation: https://scikit-learn.org/
3) Streamlit Documentation: https://streamlit.io/