

Project Description: Hand Gesture Classification Using MediaPipe Landmarks from the HaGRID Dataset

Overview:

In this project, students will work on classifying hand gestures using landmark data generated by **MediaPipe** from the **HaGRID (Hand Gesture Recognition Image Dataset)**.

The input to the project is a **CSV file** containing hand landmarks (e.g., x, y, z coordinates of keypoints) extracted from the HaGRID dataset using MediaPipe. The output will be a trained machine learning model capable of classifying hand gestures into predefined classes.

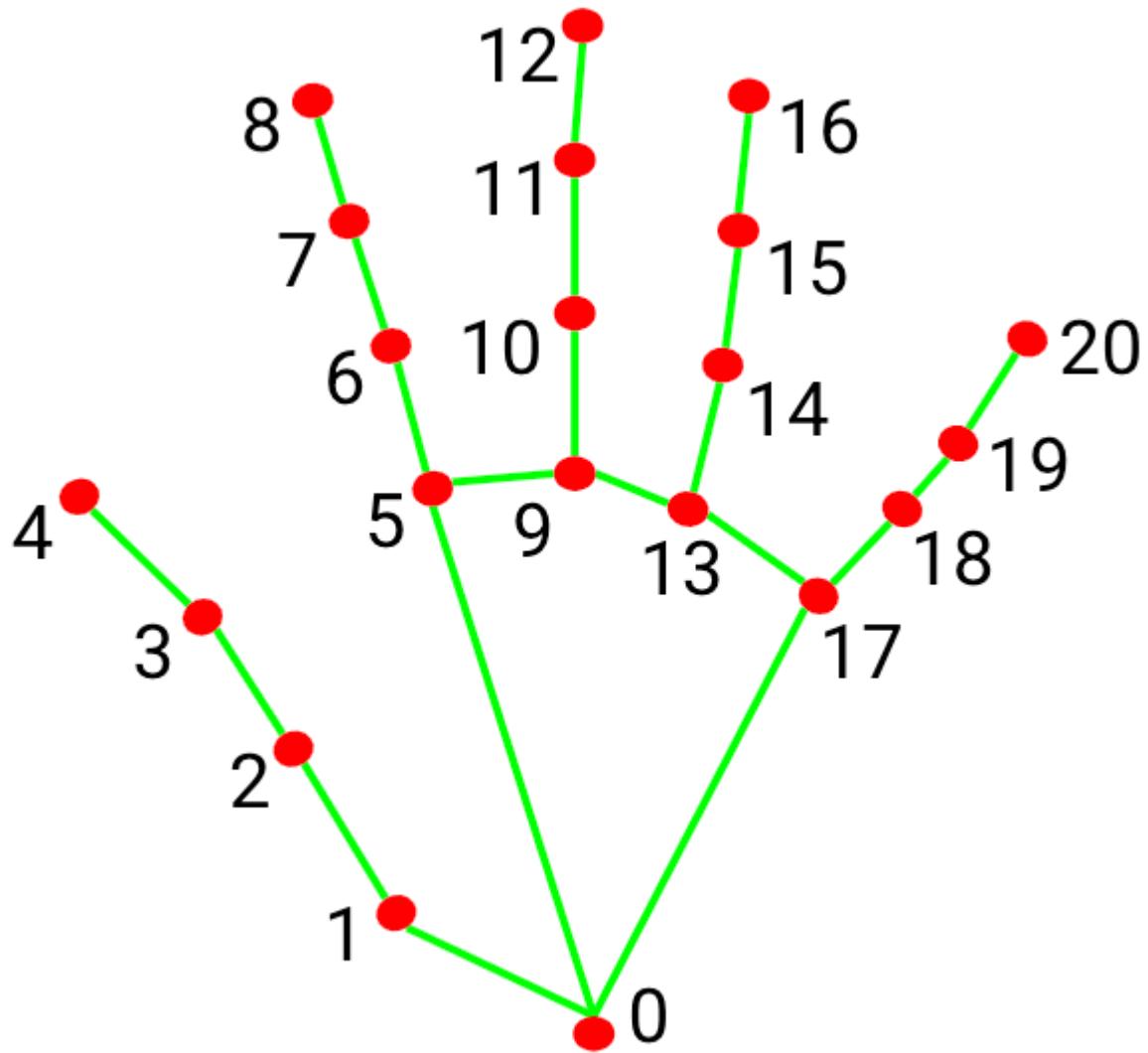
Students will gain hands-on experience in data preprocessing, visualization, machine learning model training, and performance evaluation. They will also learn how to document their work in a reproducible format and present their results effectively.

Dataset Details:

The **HaGRID dataset** contains 18 classes of hand gestures, including:



Each gesture is represented by a set of hand landmarks (21 landmarks per hand) extracted using MediaPipe. The CSV file will contain these landmarks(x,y,z location) along with their corresponding gesture labels.



Project Deliverables:

Students are required to deliver the following:

1. Google Colab Notebook:

- A well-documented and executable Colab notebook uploaded to GitHub.

- The notebook should include:
- **Data Loading:** Loading the CSV file containing hand landmarks and labels.
- **Data Visualization:** Visualizing the hand landmarks (e.g., plotting keypoints for a few samples).
- **Data Preprocessing:** Cleaning, normalizing, and preparing the data for training (e.g., handling missing values, splitting into train/test sets).
- **Model Training:** Implementing and comparing at least **3 machine learning models** (e.g., Random Forest, SVM) for gesture classification.
- **Evaluation:** Reporting accuracy, precision, recall, and F1-score for each model.
- **Conclusion:** Summarizing the results and selecting the best-performing model.

2. Output Video:

You will find a video sample in the project folder

- A short video demonstrating the results of the trained model.
- Upload the video to google drive and share a link
- You need to use **mediapipe** to get the hands landmarks in each frame

Evaluation Criteria:

1. **Code Quality:** Clean, well-documented, and reproducible code.
2. **Model Performance:** Accuracy and robustness of the trained models.
3. **Visualization:** Clarity and effectiveness of data and results visualization.

Notes:

- The detected hands have different scales and positions in the image. To overcome this problem recenter the hand landmarks (x,y) to make the origin the wrist point and divide all the landmarks by the mid-finger tip position.

With that all the detected hands will be similar in training and testing data.

- z location doesn't need to be processed as it is already processed.
- You can stabilize the output of the video by taking the mode of the outputs over a window
- Before sharing the github and video links with me, be sure that they are public and that I will be able to access them.

Good luck! 