**Introduction**

The dataset utilized in this research originates from Kaggle, ensuring a standardized format. With 3000 images per class, the dataset is balanced, each image having dimensions of 200x200x3. Preprocessing steps were employed to optimize the data for model training and real-time inference.

A collage of hands with different gestures

Description automatically generated

**A purple and white graph

Description automatically generatedA graph with purple lines

Description automatically generated**

**Preprocessing**

For the training set, images were resized to 128x128x3 to reduce computational complexity while preserving essential information. Normalization was applied to standardize pixel values between 0 and 1, enhancing data integrity and reducing redundancy.

Live images captured from a camera underwent preprocessing steps tailored for real-time hand sign classification. Utilizing Mediapipe's HandTrackingModule, the hand region was extracted from each frame. Additionally, hand contours were employed to create masks for background subtraction, ensuring robustness to varying environments. Resultant images were resized to 128x128x3, and pixel values rescaled to match the training set.

**Model Training**

Various convolutional neural network (CNN) architectures were explored to identify the most effective model.

A close up of a white sheet

Description automatically generated with medium confidence**a) Customized CNN with large no. of hidden layers:**

The accuracy of the model for testing data is: 3.703

The loss of the model for testing data is: 3.295.

Correctly predicted classes: 450

Incorrectly predicted classes: 11700

A graph of a graph

Description automatically generated with medium confidence

A graph of a graph

Description automatically generated with medium confidence

A customized CNN with an extensive number of hidden layers exhibited overfitting, rendering it unsuitable for further consideration.

A diagram of a company

Description automatically generated with medium confidence**b) Customized CNN with few hidden layers:**

The accuracy of the model for testing data is: 99.934

The loss of the model for testing data is: 0.0023

Correctly predicted classes: 12142

A graph of loss and loss

Description automatically generatedA graph of loss and loss

Description automatically generatedIncorrectly predicted classes: 8

Customized CNN with a reduced number of Conv2D layers demonstrated exceptional performance on the testing dataset, achieving an accuracy of 99.93% and minimal loss. This model correctly predicted 12,142 out of 12,150 test images.

A graph of loss and loss

Description automatically generatedA graph of loss and loss

Description automatically generatedA close up of a paper

Description automatically generated**c) VGG16 with ImageNet weights:**

The accuracy of the model for testing data is: 99.341

The loss of the model for testing data is: 0.035  
Correctly predicted classes: 12070

Incorrectly predicted classes: 80

VGG16, a widely recognized CNN architecture, was utilized with pretrained weights from ImageNet. VGG16 consists of 13 convolutional layers and 3 fully connected layers. The convolutional layers in VGG-16 are all 3×3 convolutional layers with a stride size of 1 and the same padding, and the pooling layers are all 2×2 pooling layers with a stride size of 2. VGG16 is renowned for its simplicity and effectiveness in image classification tasks. Despite achieving a slightly lower accuracy of 99.34% on the testing dataset, VGG16 exhibited superior performance when presented with real-time images, accurately classifying hand signs.

**A screen shot of a chart

Description automatically generatedFeature Extraction & Explainability using t-SNE graph from VGG16 model:**

The t-SNE, or t-Distributed Stochastic Neighbor Embedding, is a dimensionality reduction technique commonly used in machine learning and data visualization tasks. In hand sign recognition, t-SNE is utilized to visualize the distribution and clustering of hand sign images in a lower-dimensional space. First, features are extracted from the images using VGG16 architecture of CNN. t-SNE then reduces the dimensionality of these features while preserving their local structure. The resulting 2D scatter plot allows for the identification of patterns and clusters corresponding to different hand signs, aiding in dataset analysis and comparison across conditions or datasets. From the graph, it is evident that most of the images are correctly classified by the VGG16 model with a few exceptions.

**Conclusion**

In conclusion, while the customized CNN with fewer Conv2D layers demonstrated outstanding performance on the testing dataset, the VGG16 model pretrained on ImageNet weights emerged as the optimal choice for real-time American Sign Language classification. Its robustness and accuracy in handling real-world data underscore its suitability for practical applications in sign language recognition systems.