DA-3
SIGN LANGUAGE RECOGNITION USING DEEP LEARNING AND IMAGE PROCESSING

MANOJ KUAMAR P

21BAI1040

INTRODUCTION

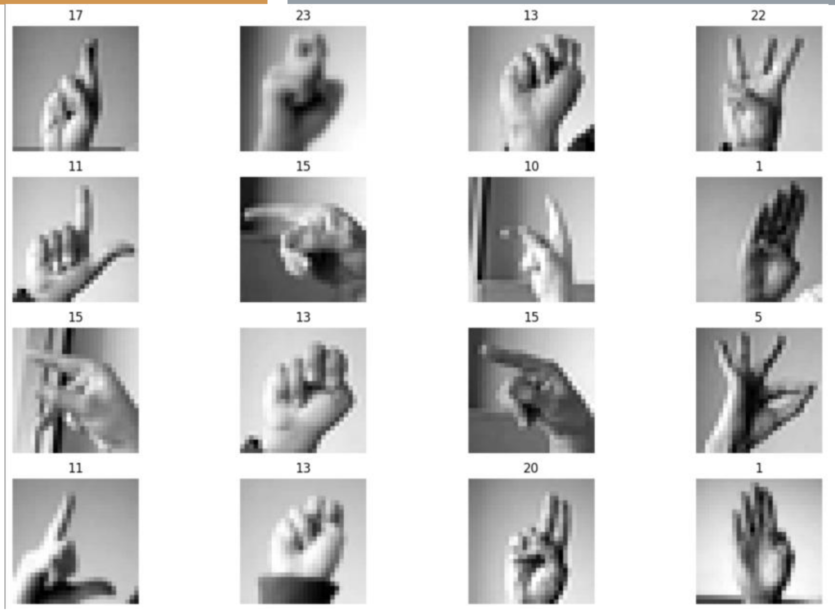
- MNIST dataset served as a benchmark for machine learning.
- Fashion-MNIST dataset introduced to address limitations of MNIST.
- Sign Language MNIST dataset aims to provide a more challenging task for computer vision.
- Inspired by Fashion-MNIST, Sign Language MNIST follows a similar CSV format.
- Represents a multi-class problem with 24 classes of letters in American Sign Language (ASL).

OBJECTIVES

- Train a deep learning model to recognize hand gestures representing letters in ASL.
- Evaluate the model's performance using metrics like precision, recall, and F1 score.
- Develop a real-world application for sign language recognition using computer vision techniques.

DATASET AND ATTRIBUTES

- **Sign Language MNIST:**
  - Format closely mimics classic MNIST dataset.
  - Consists of 27,455 training cases and 7,172 test cases.
  - Each case represents a label (0-25) corresponding to an alphabetic letter A-Z.
  - No cases for 9=J or 25=Z due to gesture motions.
  - Each image is 28x28 pixels with grayscale values between 0-255.
  - Data augmentation techniques used to expand the dataset, including cropping, resizing, rotation, and pixelation.




```
label,pixel1,pixel2,pixel3,pixel4,pixel5,pixel6,pixel7,pixel8,pixel9,pixel10,pixel11,pixel12,pixel13,pixel14,pixel15,pixel16,pixel17,pixel18,pixel19,pixel
6,149,149,150,150,150,151,151,150,151,152,152,152,152,152,153,153,151,152,152,153,152,152,151,151,150,150,150,149,150,150,150,152,152,151,152,152,152,152,
5,126,128,131,132,133,134,135,135,136,138,137,137,138,138,139,137,142,140,138,139,137,137,136,135,134,133,134,132,129,132,134,135,135,137,139,139,139,140,
10,85,88,92,96,105,123,135,143,147,152,157,163,168,171,182,172,175,185,183,184,185,185,185,183,183,182,181,178,86,88,93,96,108,125,137,145,149,154,160,166
0,203,205,207,206,207,209,210,209,210,209,208,207,207,209,208,210,210,207,209,209,208,209,210,209,207,208,209,207,206,208,209,208,208,210,211,210,211,209,
3,188,191,193,195,199,201,202,203,203,203,204,204,204,203,202,198,216,217,135,181,200,195,194,193,190,189,187,185,190,194,196,197,200,202,204,206,207,207,
21,72,79,87,101,115,124,131,135,139,142,144,147,150,153,156,159,160,162,164,165,166,166,167,167,168,168,168,167,73,80,89,104,117,126,132,136,140,143,146,1
10,93,100,112,118,123,127,131,133,136,139,140,143,144,145,146,149,151,153,154,155,156,159,159,160,160,161,163,164,93,102,113,119,123,128,131,134,138,140,1
14,177,177,177,177,177,178,179,179,178,179,179,178,179,178,179,179,179,179,178,178,179,178,177,178,177,177,179,143,180,180,180,180,180,180,181,182,183,182
3,191,194,196,198,201,203,204,205,205,205,205,206,206,205,204,200,217,218,143,185,202,198,197,195,193,192,190,189,193,197,198,200,203,204,206,208,208,208,
7,171,172,172,173,173,173,173,173,172,172,172,171,170,170,169,168,168,166,165,165,164,164,152,86,72,61,65,85,174,174,174,175,175,175,175,175,175,174,174,1
8,212,212,213,212,214,213,213,213,214,215,214,213,212,213,212,212,211,211,212,211,210,210,207,207,207,210,191,83,213,214,215,216,216,214,215,216,216,214,2
8,187,186,187,186,188,187,187,187,188,188,188,187,186,187,187,186,186,186,186,185,185,184,182,181,182,185,168,75,187,188,189,189,189,188,189,190,190,188,1
21,128,131,133,135,137,139,140,142,145,146,146,145,144,149,147,147,148,148,146,147,147,147,149,148,145,145,144,142,129,132,134,137,139,140,143,145,146,147
12,178,179,181,183,184,183,183,184,185,186,185,184,184,183,185,187,182,179,179,177,177,177,174,173,172,170,169,166,179,184,185,185,186,187,186,186,187,187
7,119,120,120,121,120,120,120,119,119,120,120,120,121,120,119,118,118,118,109,32,16,37,51,52,71,91,102,121,121,122,123,123,123,122,122,122,123,122,123,122
4,191,192,192,192,192,193,193,193,193,192,193,193,192,192,192,191,190,187,188,188,187,186,184,184,182,182,181,180,193,194,195,195,195,195,195,195,195,194,
22,31,59,78,82,86,90,94,105,119,129,136,139,142,146,149,153,156,158,159,160,162,163,164,164,163,164,163,162,31,62,79,83,87,91,95,106,120,131,137,140,143,1
0,199,201,203,205,207,208,209,211,211,212,212,212,212,212,212,212,212,211,211,211,211,210,210,210,211,210,208,207,200,201,204,206,208,210,211,212,213,213,
7,95,105,119,144,155,163,169,172,173,177,179,181,182,183,184,185,186,185,184,182,181,181,180,178,176,175,172,170,95,106,121,145,157,164,170,173,175,179,18
7,121,131,138,138,143,149,156,159,162,164,167,169,171,173,175,177,178,179,179,179,178,179,179,178,177,177,176,175,122,132,137,139,144,150,156,160,163,166,
```

PROPOSED METHODOLOGY

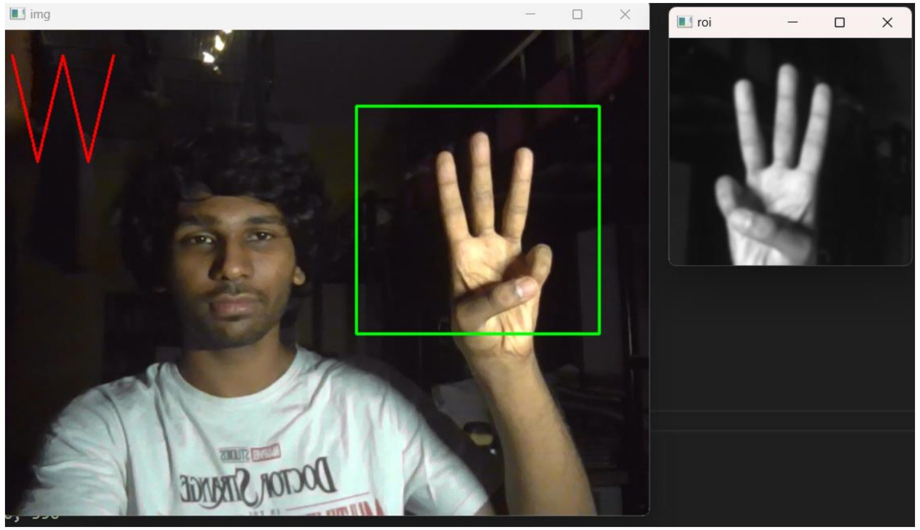
Proposed Methodology:

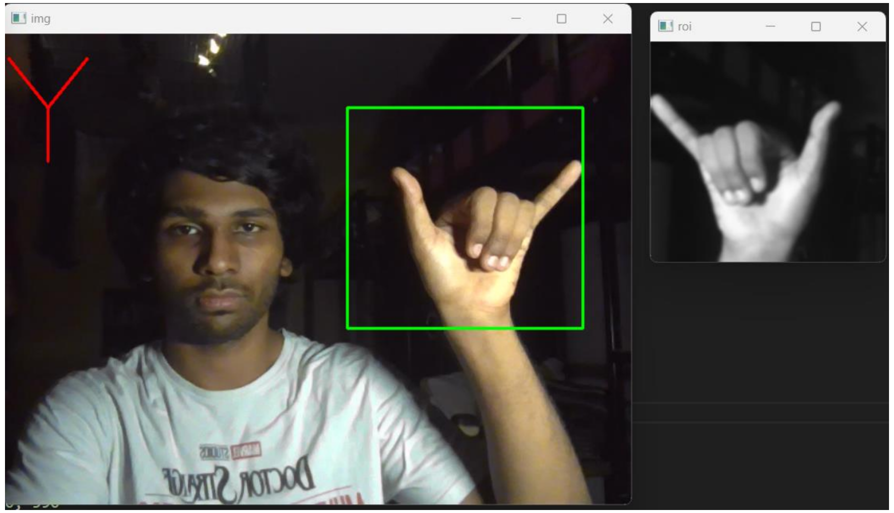
- **Custom CNN Architecture:**
  1. Designing a tailored CNN architecture for sign language gestures.
  2. Experimenting with layers to optimize feature extraction and model complexity.
- **Data Augmentation:**
  1. Applying preprocessing (cropping, resizing, grayscale) for standardization.
  2. Employing augmentation (rotation, zoom, shifts) to increase dataset diversity.
- **Optimization Algorithms:**
  1. Using RMSprop, Adam, etc., for efficient loss minimization.
  2. Employing early stopping, learning rate adjustments for dynamic training.
- **Model Evaluation:**
  1. Assessing performance with accuracy, precision, recall, F1 score.
  2. Validating on training/test datasets for generalization.
- **Interpretability:**
  1. Visualizing CNN activations, feature maps for insights.
  2. Analyzing misclassifications for model refinement.
- **Interactive User Interface:**
  1. Creating an interactive application for real-time gesture recognition.
  2. Enabling users to communicate through sign language gestures in an intuitive manner.
- **Evaluation and User Feedback:**
  1. Collecting user feedback to assess the effectiveness and usability of the application.
  2. Evaluating model performance and user experience for further refinement.

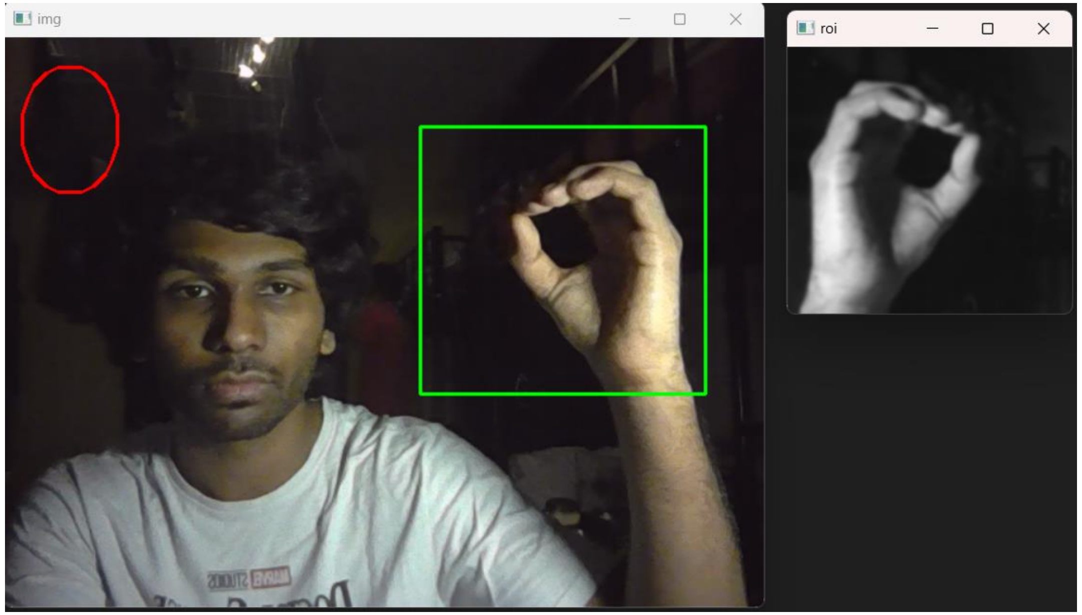
```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 28, 28, 64)        1664

 conv2d_1 (Conv2D)           (None, 28, 28, 64)        102464

 max_pooling2d (MaxPooling2  (None, 14, 14, 64)        0
 D)

 dropout (Dropout)           (None, 14, 14, 64)        0

 conv2d_2 (Conv2D)           (None, 14, 14, 32)        18464

 conv2d_3 (Conv2D)           (None, 14, 14, 32)        9248

 max_pooling2d_1 (MaxPoolin  (None, 7, 7, 32)          0
 g2D)

 dropout_1 (Dropout)         (None, 7, 7, 32)          0

 flatten (Flatten)           (None, 1568)              0

 dense (Dense)               (None, 256)               401664
...
Total params: 539929 (2.06 MB)
Trainable params: 539929 (2.06 MB)
Non-trainable params: 0 (0.00 Byte)
```

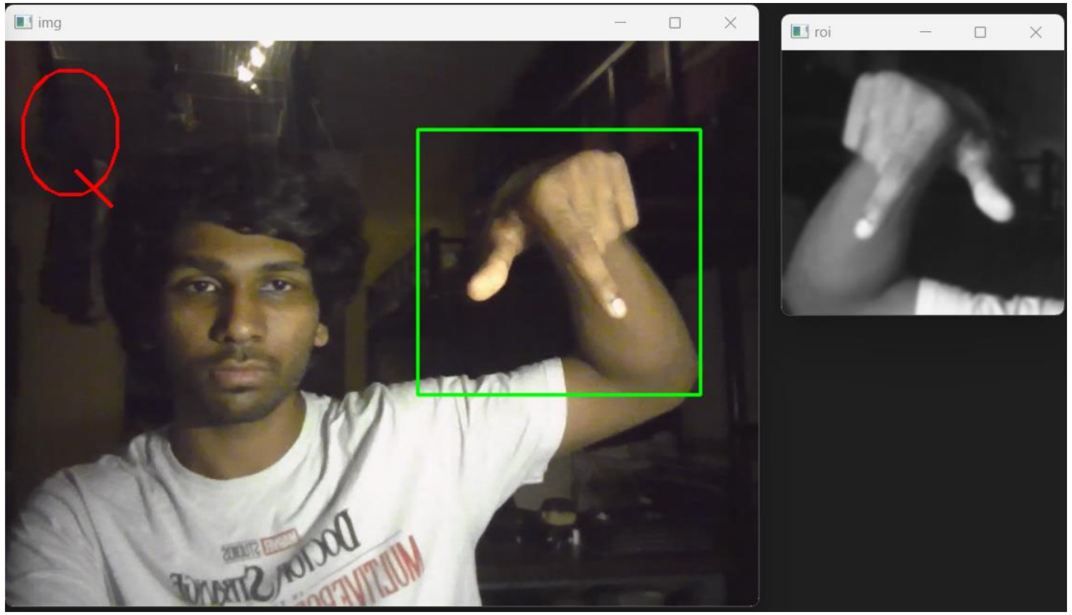
EXPERIMENTAL RESULTS

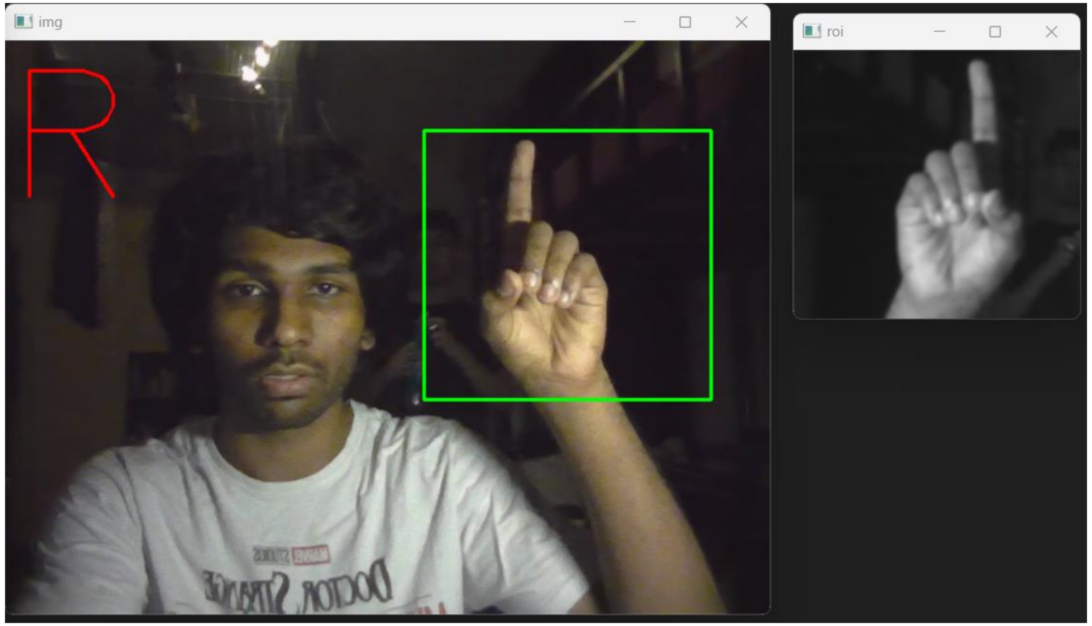
- **Camera Calibration:** Prior to gesture capture, camera calibration is performed to ensure accurate depth perception and spatial mapping of hand gestures.
- **Real-time Processing:** The system employs efficient algorithms for real-time processing of captured frames, minimizing latency and ensuring smooth interaction.
- **Cross-Platform Compatibility:** The software is designed to be compatible with multiple platforms, including Windows, macOS, and Linux, enhancing its accessibility and versatility.
- **User Interface Design:** A user-friendly interface is developed to provide intuitive controls and visual feedback, enhancing the user experience during gesture recognition.

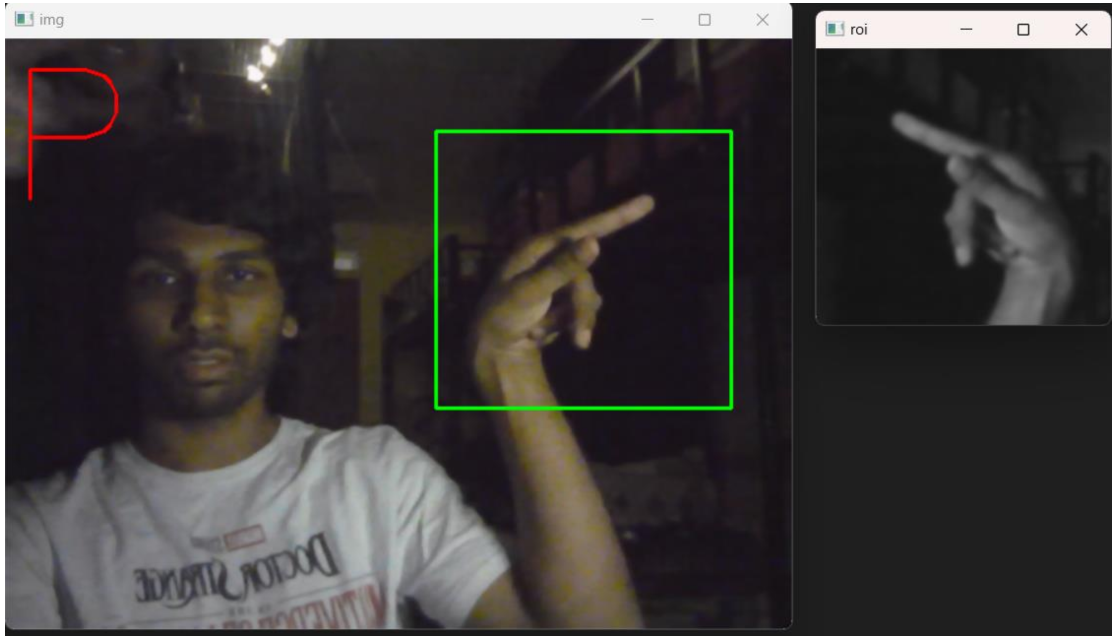





img
roi
O


img
roi
Q


img
roi
R


img
roi
P

PERFORMANCE EVALUATION

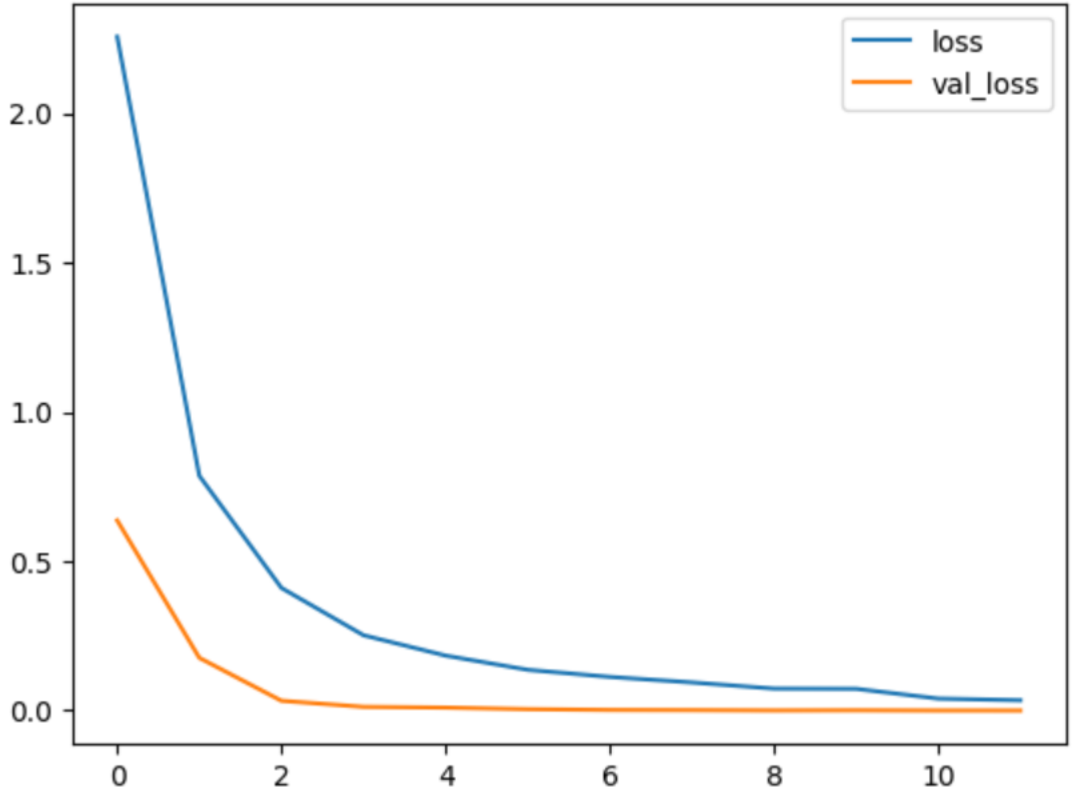
- **Model Tuning:** The project explores various hyperparameters and optimization techniques to fine-tune the CNN architecture for optimal performance.
- **Validation Strategies:** Cross-validation techniques such as k-fold validation are utilized to ensure robustness and reliability of the model's performance evaluation.
- **Scalability Assessment:** The scalability of the system is evaluated to determine its capability to handle increased workload and user interactions without compromising performance.
- **Error Analysis:** Detailed error analysis is conducted to identify common misclassifications and areas for model improvement, guiding future iterations and refinements.

```
57/57 [==============================] - 1s 8ms/step - loss: 0.0211 - accuracy: 0.9916
Testing accuracy of model : 99.16%
```

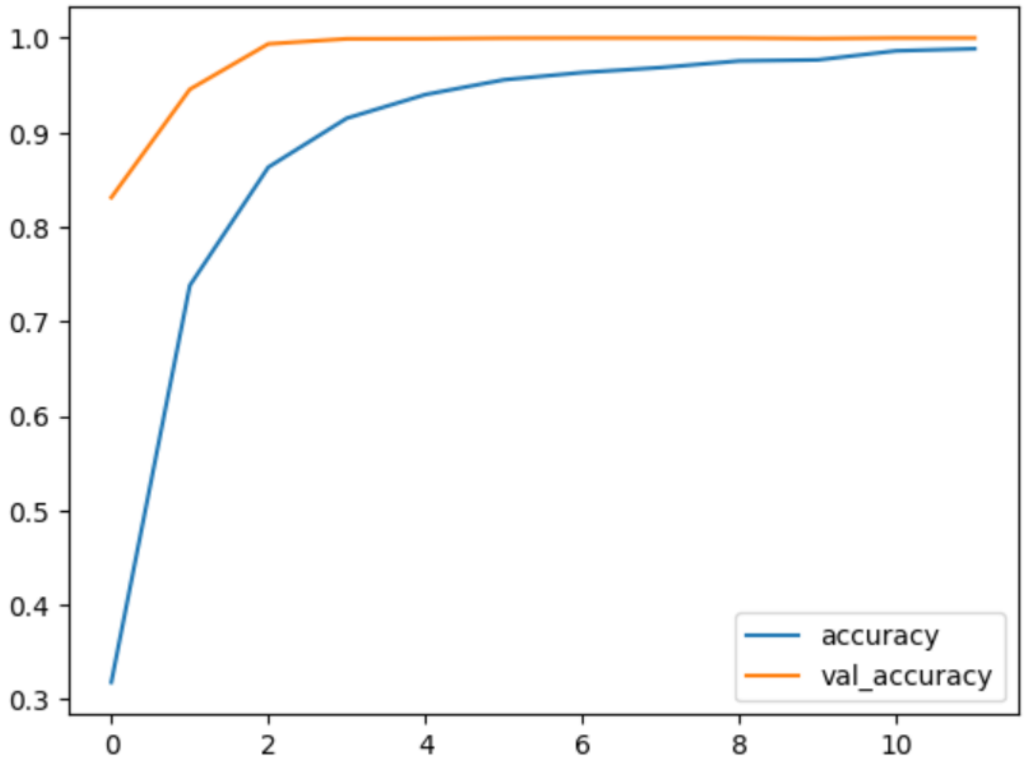
```
225/225 [==============================] - 0s 2ms/step
F1 Score: 0.9537261059762697
```

```
Label: 0
Precision: 1.0
Recall: 1.0
F1 Score: 1.0

Label: 1
Precision: 1.0
Recall: 1.0
F1 Score: 1.0

Label: 2
Precision: 0.9841269841269841
Recall: 1.0
F1 Score: 0.992

Label: 3
Precision: 1.0
Recall: 1.0
F1 Score: 1.0

Label: 4
Precision: 1.0
Recall: 1.0
F1 Score: 1.0

Label: 5
Precision: 1.0
Recall: 1.0
F1 Score: 1.0
```

```
Label: 6
Precision: 0.9935691318327974
Recall: 0.8879310344827587
F1 Score: 0.9377845220030349

Label: 7
Precision: 0.956140350877193
Recall: 1.0
F1 Score: 0.9775784753363229

Label: 8
Precision: 0.9442622950819672
Recall: 1.0
F1 Score: 0.9713322091062394

Label: 9
Precision: 0.0
Recall: 0.0
F1 Score: 0.0

Label: 10
Precision: 1.0
Recall: 1.0
F1 Score: 1.0

Label: 11
Precision: 1.0
Recall: 1.0
F1 Score: 1.0
```



loss
val_loss
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
10



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
2
4
6
8
10
accuracy
val_accuracy

RELATED WORKS AND LIMITATIONS

- **Comparison:** Existing sign language recognition systems often require complex setups and specialized hardware, limiting their scalability and practicality. While some systems achieve high accuracy, they may lack real-time performance or require extensive training data.
- **Pros and Cons:** Prior approaches have demonstrated advancements in gesture recognition accuracy, but they often face challenges related to cost, hardware requirements, and real-world deployment.
- **Citations:**

  - *Wadhawan, A., Kumar, P. Deep learning-based sign language recognition system for static signs. Neural Comput & Applic 32, 7957–7968 (2020). https://doi.org/10.1007/s00521-019-04691-y*
  - *N. Adaloglou et al., "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," in IEEE Transactions on Multimedia, vol. 24, pp. 1750-1762, 2022, doi: 10.1109/TMM.2021.3070438*
  - *https://www.kaggle.com/datasets/datamunge/sign-language-mnist*

CONCLUSION AND FUTURE WORK

- **Conclusion:** In conclusion, the project demonstrates significant progress in real-time sign language recognition, contributing to accessibility and inclusivity for the deaf and hard-of-hearing community.
- **Future Directions:** Future research directions include investigating multimodal approaches incorporating audio and text-based inputs, enhancing the system's robustness in challenging environments, and expanding its functionality to support additional sign languages and communication modalities.
- **Community Engagement:** Engaging with the deaf and hard-of-hearing community in co-designing and co-developing future iterations of the system is essential to ensure its relevance, effectiveness, and inclusivity.
- **Long-term Impact:** The long-term impact of the project on society, education, and healthcare is considered, emphasizing the importance of leveraging technology for positive social change and empowerment.

THANK YOU