American Sign Language Recognition using Convolutional Neural Networks

Fatima-Zahrae El-Qoraychy, Yazan Mualla

CIAD, University of Technology of Belfort-Montbéliard (UTBM)

fatima.el-qoraychy@utbm.fr
yazan.mualla@utbm.fr

Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services
CENTRIC 2023
Valencia, Spain, November 13-17, 2023

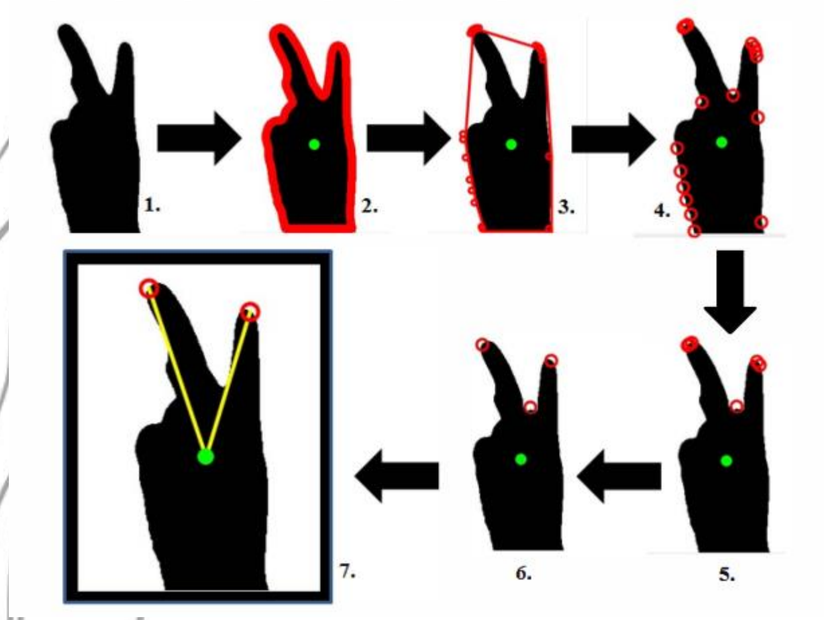
Presenter

- Name: Fatima-Zahrae El-Qoraychy
- Degree: Master's in Data Science and Artificial Intelligence
- University: University of Sorbonne Paris Nord, France
- Year of Degree: 2023
- Doctoral Studies: Currently pursuing a Ph.D. at the CIAD Laboratory, University of UBFC
- Research Area: Socially Smart Connected and Autonomous Vehicles
- Research Focus: Enhancing the safety and efficiency of cooperative intersections using Deep Reinforcement Learning

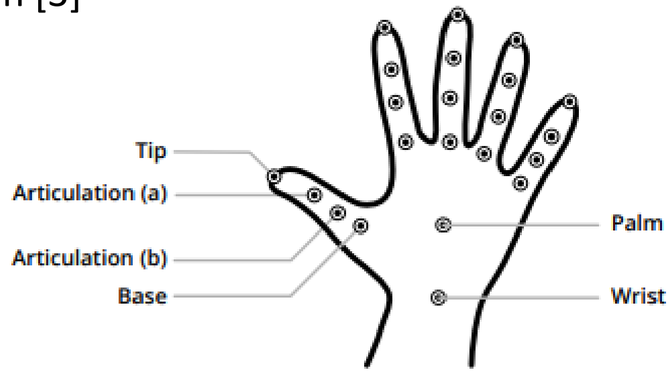
Problematic

How can we use artificial intelligence to predict and interpret hand gestures in sign language to facilitate communication and interaction between deaf and hard-of-hearing individuals and their environment?

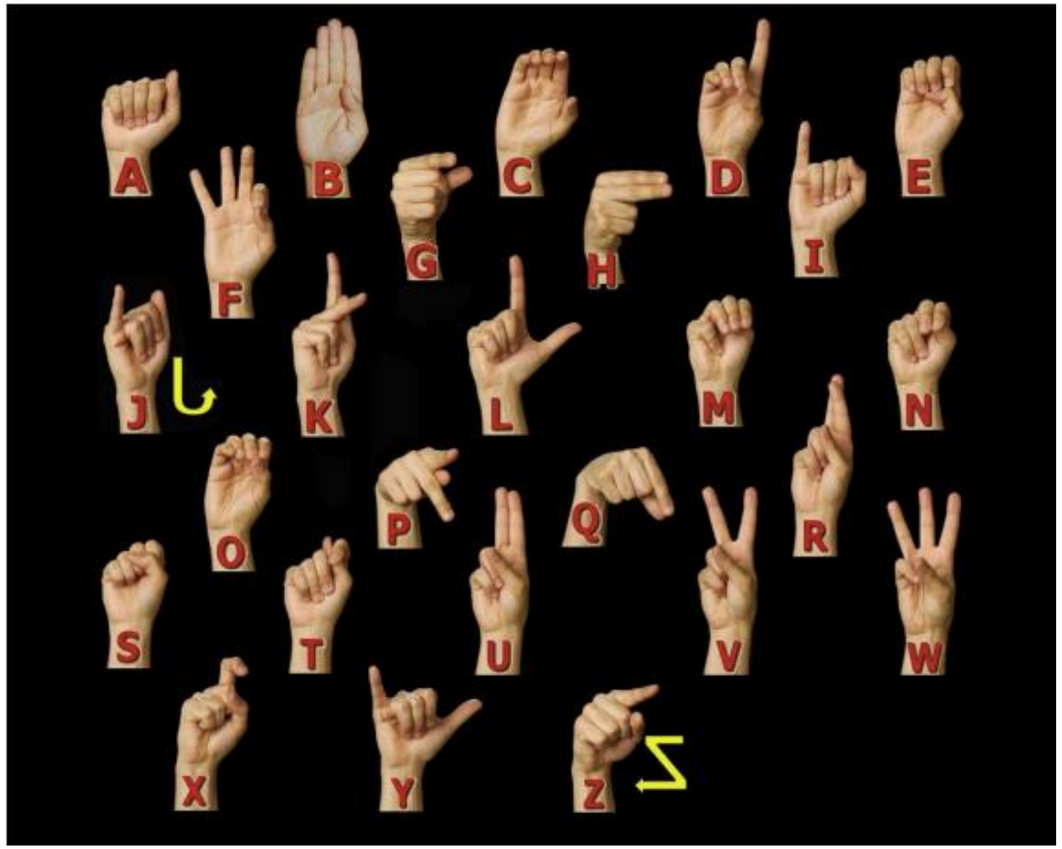
Advances in Sign Language Recognition



Finger-tip detection [3]



Hand skeleton [4]



American sign language finger spelling alphabet [5]

Existing Solution

kaggle

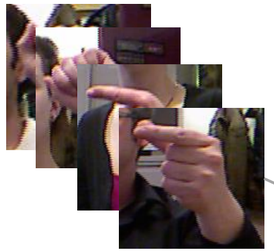
Link [6]

DAMION JOYNER · COPIED FROM BARRY +352,-100 · 1Y AGO · 432 VIEWS

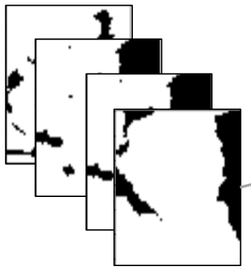
Sign-Language-Classification-CNN-VGG19

Python · VGG-19, ASL Fingerspelling Images (RGB & Depth), [Private Datasource]

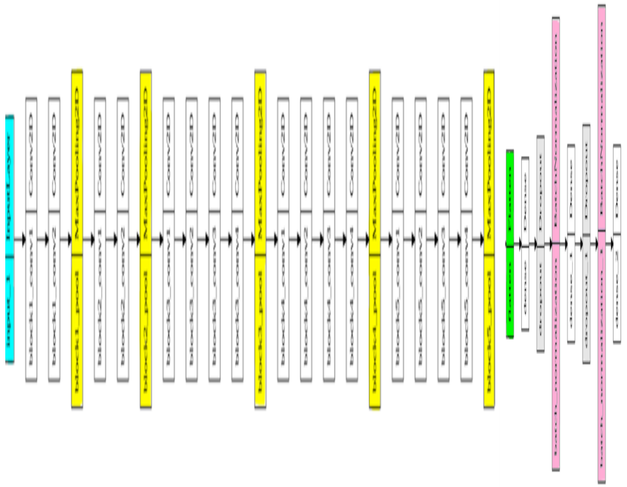
Existing Method Structure

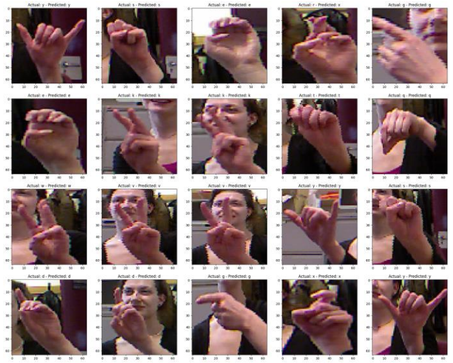


RGB



Depth





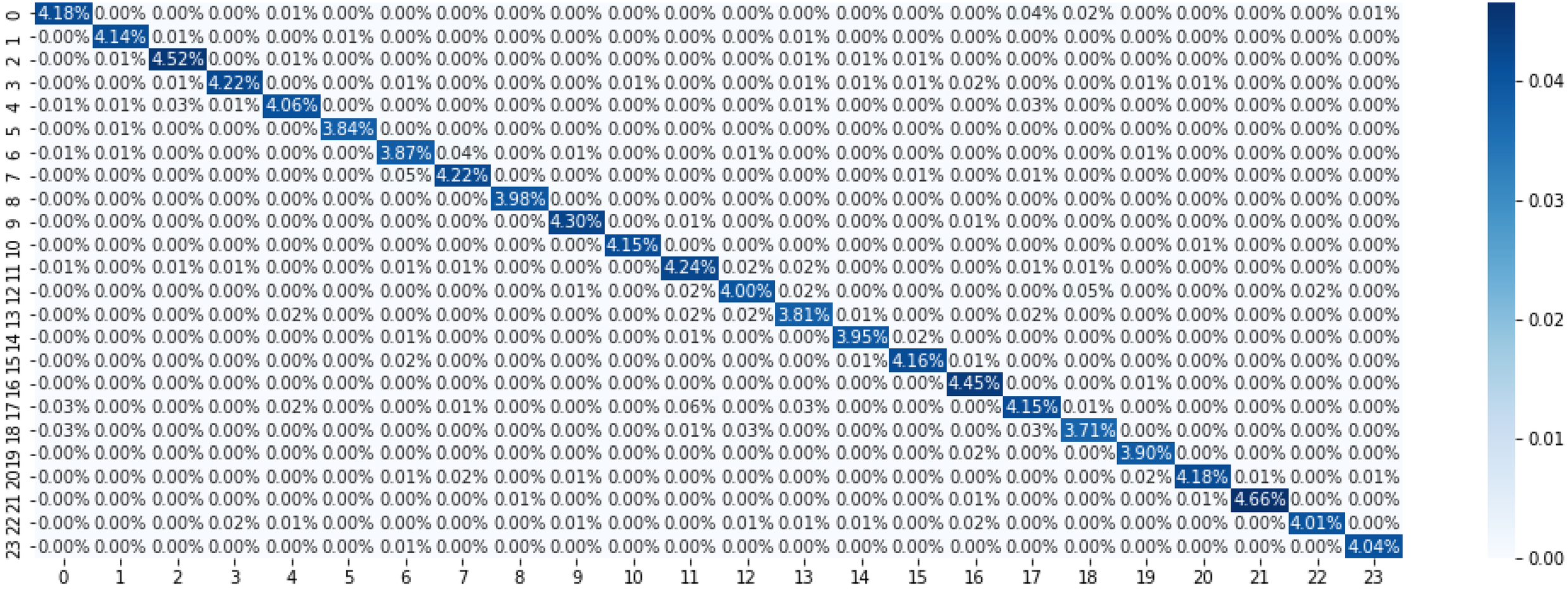
100 000 Labeled images [7]

Classification model

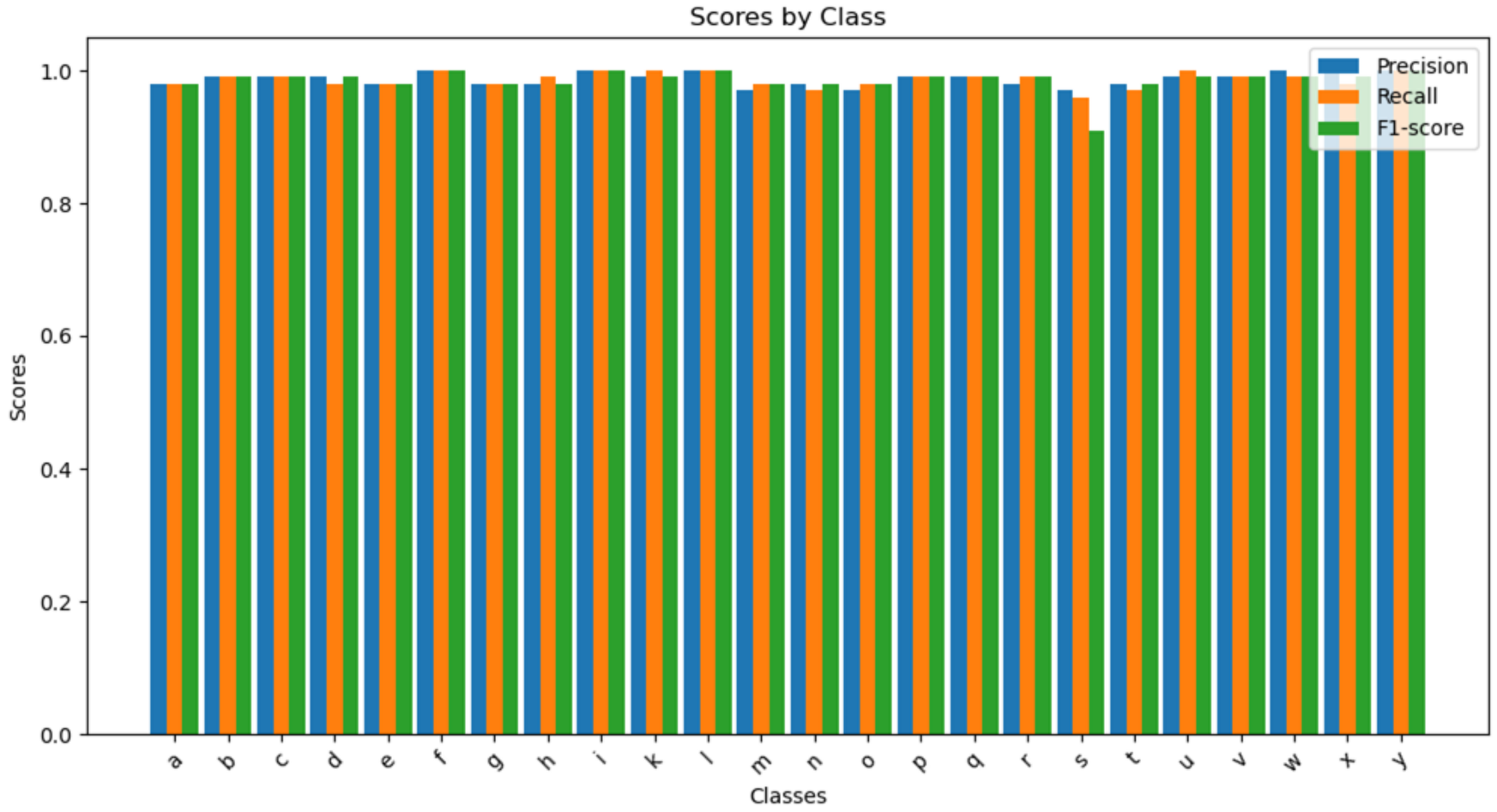
Prediction

Existing Method Performance Analysis

Confusion matrices

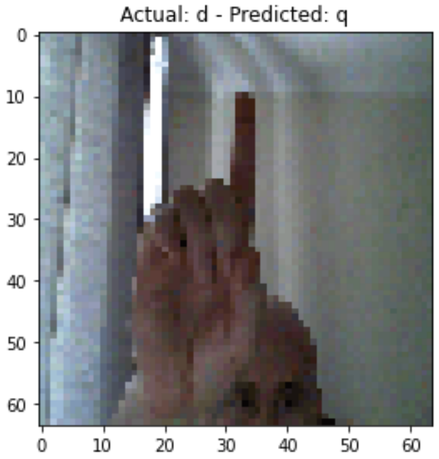


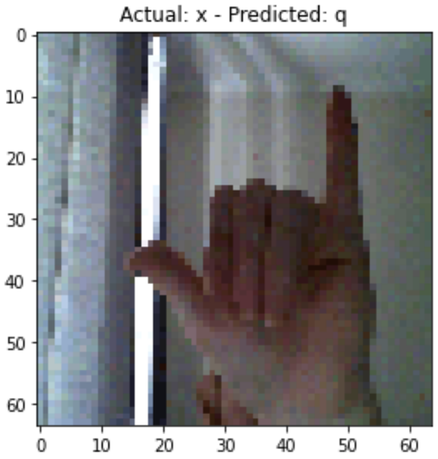
Existing Method Performance Analysis

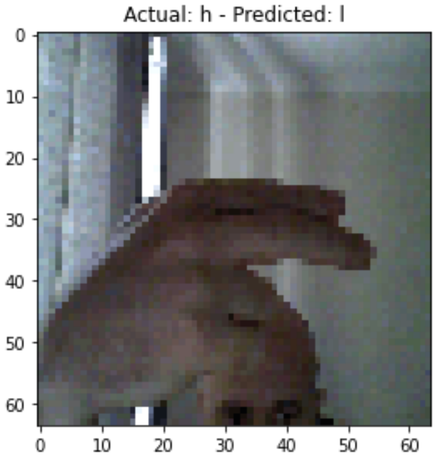


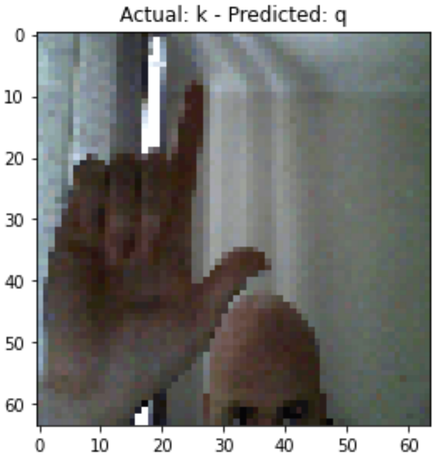
Existing Method Test Results

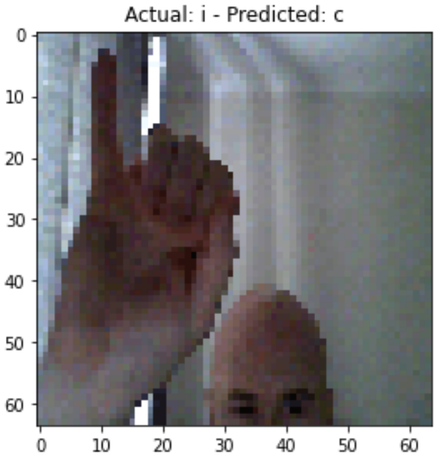
- The test results

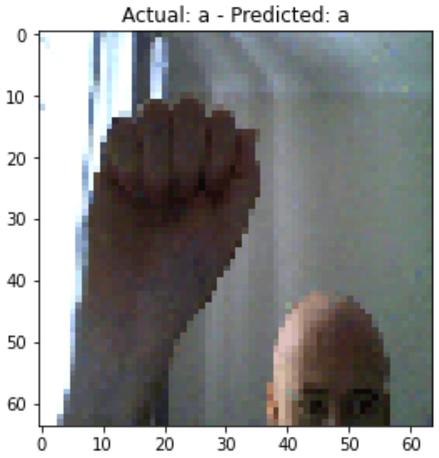












Method Limitations and Proposed Enhancements

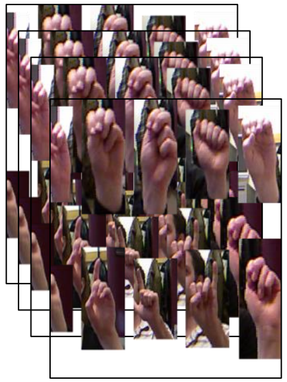
Limitations of the Classification Model

- ❑ Lack of Data Diversity
- ❑ Overfitting to Specific Hand Features

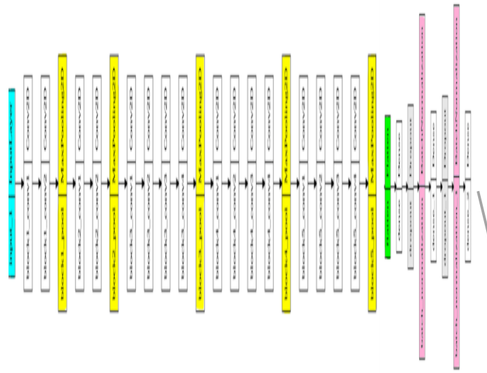
Our Proposed Enhancements

- ❑ Data Diversification
- ❑ Data Augmentation
- ❑ Using Segmented Images

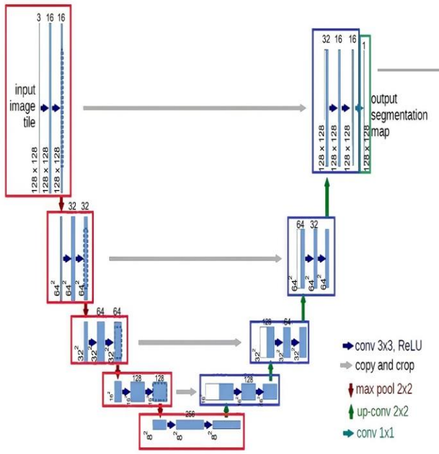
Proposed Solution: Structure



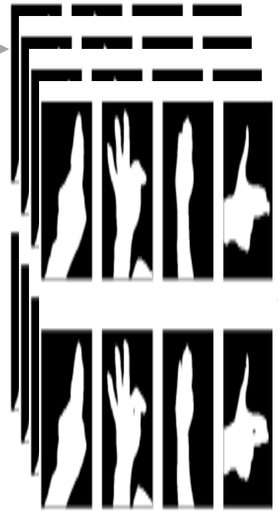
	directories	labels	images
0	A	a	A/a/color_0_0002.png
1	A	a	A/a/color_0_0003.png
2	A	a	A/a/color_0_0004.png
3	A	a	A/a/color_0_0005.png
4	A	a	A/a/color_0_0006.png



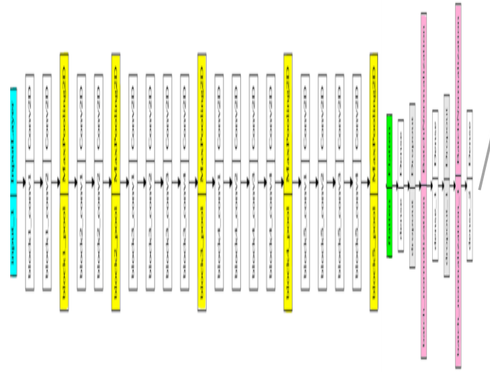
436,433 Labeled images
[7] [8] [9] [10] [11]



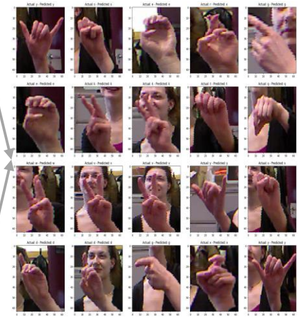
UNET



	directories	labels	images
0	A	a	A/a/color_0_0002_mask.png
1	A	a	A/a/color_0_0003_mask.png
2	A	a	A/a/color_0_0004_mask.png
3	A	a	A/a/color_0_0005_mask.png
4	A	a	A/a/color_0_0006_mask.png



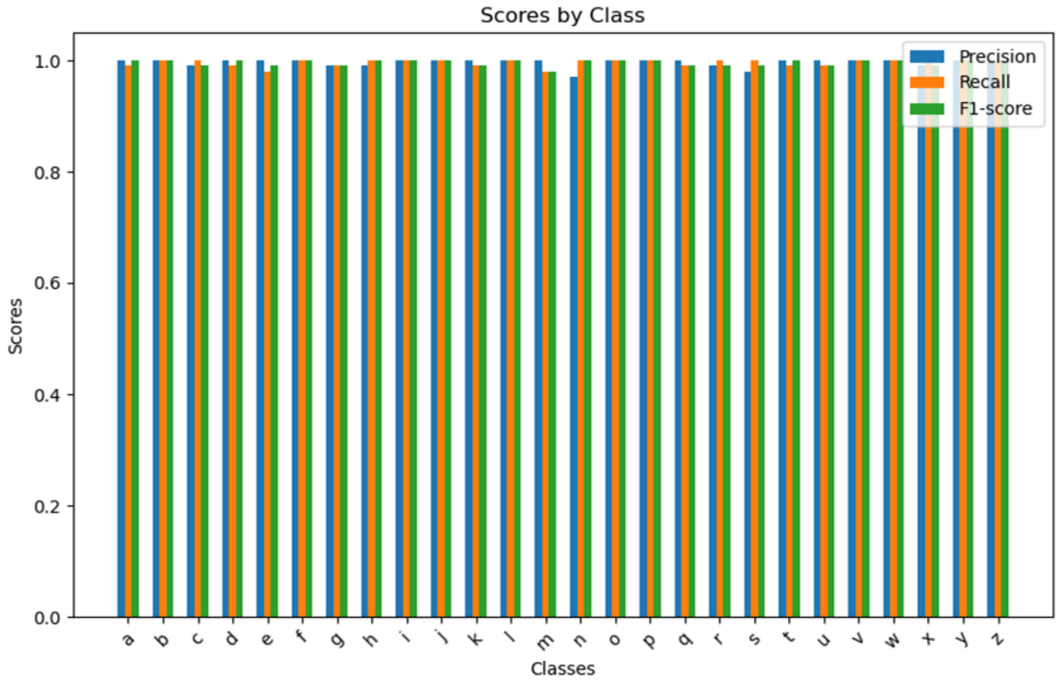
Prediction



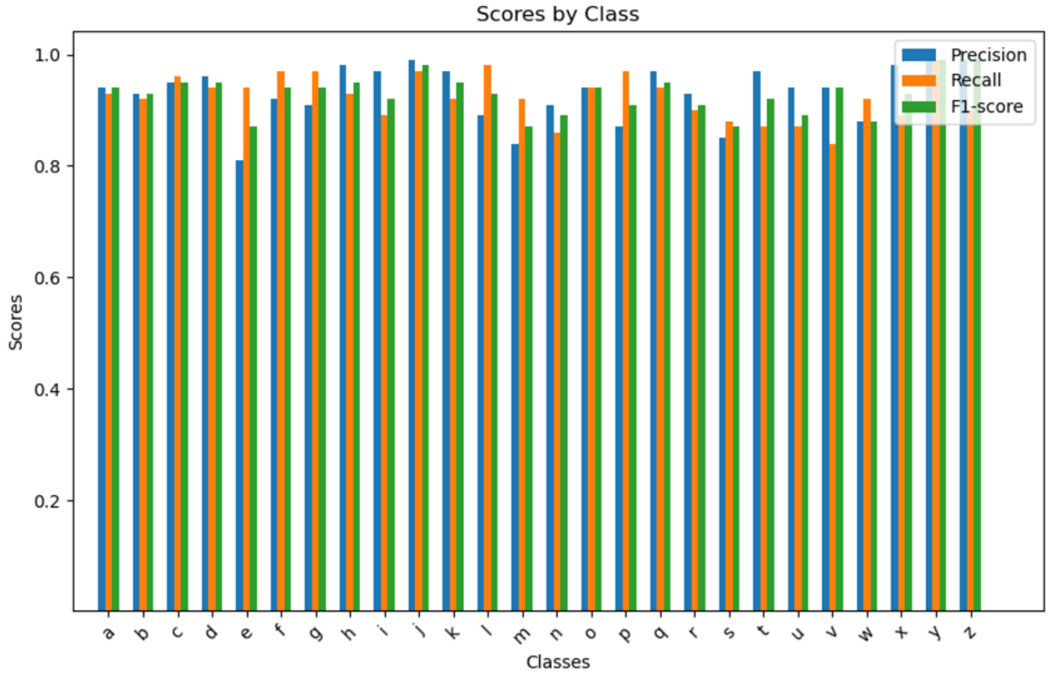
Proposed Solution: Performance Analysis

Confusion matrices

Proposed Solution: Performance Analysis



Results of the RGB model



Results of the mask model

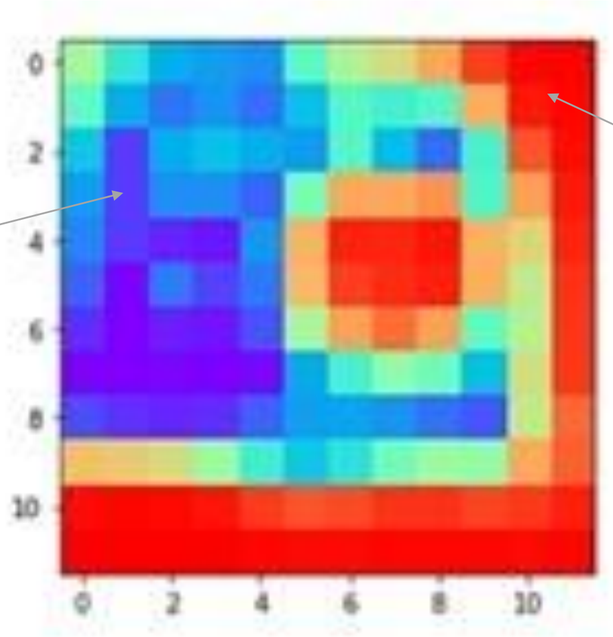
Result of models

Weighted average	Accuracy	F1 score	Precision
RGB model	0.99	0.99	0.99
Mask model	0,93	0,93	0,93

Enhancing Model Explainability with Grad-CAM

Grad-CAM (Gradient-weighted Class Activation Mapping):

- ❏ An interpretive artificial intelligence technique used to identify and illustrate noteworthy regions within an image during the classification process
- ❏ It generates a heat map that emphasizes the areas most impactful for predicting the class

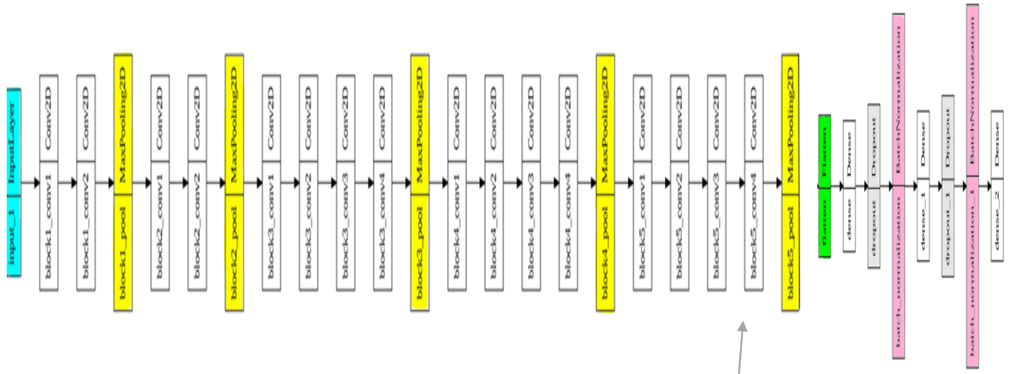


Less Important

More Important

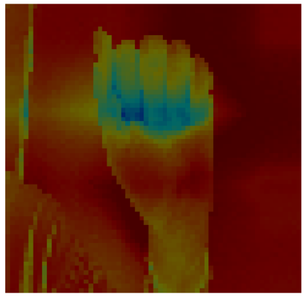
Enhancing Model Explainability with Grad-CAM

Grad-Cam on RGB Image Classification Model



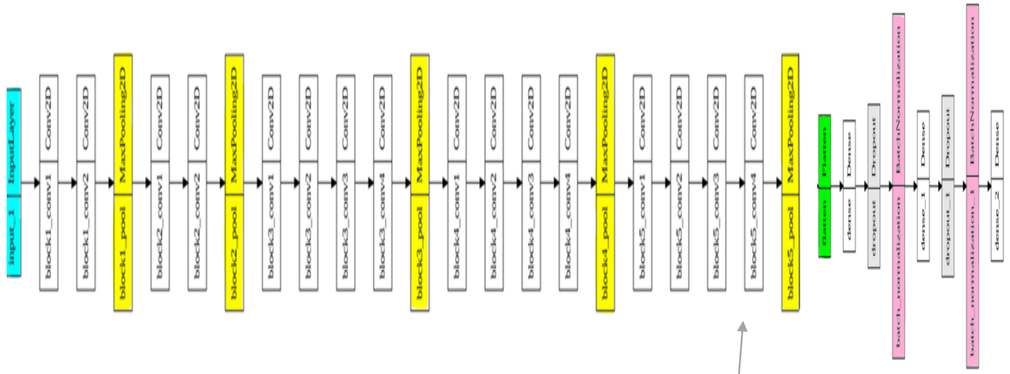
Prediction a

Grad-Cam



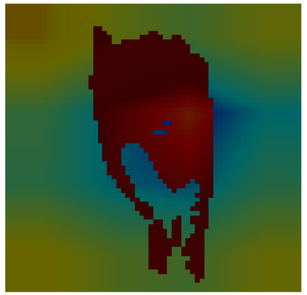
Enhancing Model Explainability with Grad-CAM

Grad-Cam on Mask Image Classification Model

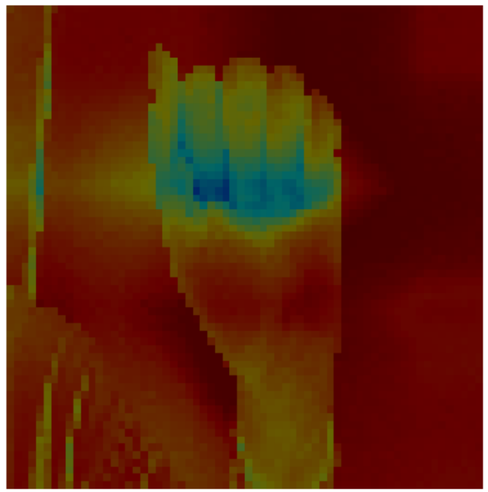


Prediction a

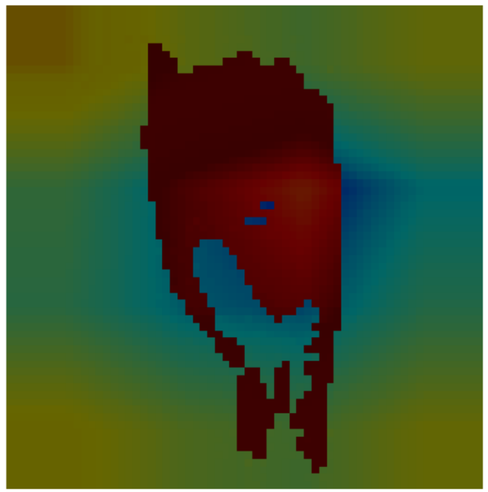
Grad-Cam



Enhancing Model Explainability with Grad-CAM



RGB model



Mask model

Conclusion and Future Work

- ❑ Extend the project to handle expressions
- ❑ Extend the work to process video and enable real-time communication with a hard of hearing or deaf person
- ❑ Development of a framework that will be used to communicate with this category of individuals
- ❑ Utilize other explainable models

References

[1] https://www.ciad-lab.fr/yazan_mualla/

[2] https://www.ciad-lab.fr/fatima-zahrae_el-goraychy/

[3] Anup Kumar, Karun Thankachan, and Mevin M. Dominic. Sign language recognition. In 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), pages 422–428, 2016.

[4] Guillaume Devineau, Wang Xi, Fabien Moutarde, and Jie Yang. Dee learning for hand gesture recognition on skeletal data. In 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), page 106–113. IEEE, 2018

[5] Brandon Garcia and Sigberto Alarcon Viesca. Real-time american sign language recognition with convolutional neural networks. volume 2, page 8. Stanford University Stanford, CA, USA, 2016.

[6] https://www.kaggle.com/code/damionjoyner/sign-language-classification-cnn-vgg19

[7] https://www.kaggle.com/datasets/mrgeislinger/asl-rgb-depth-fingerspelling-spelling-it-out

[8] https://www.kaggle.com/datasets/grassknoted/asl-alphabet

[9] https://www.kaggle.com/datasets/danrasband/asl-alphabet-test

[10] https://www.kaggle.com/datasets/debashishsau/aslamerican-sign-language-aplhabet-dataset

[11] https://www.kaggle.com/datasets/lexset/synthetic-asl-alphabet