Sign Language Classification

Literature Review

by:

Youssef Metwally

Khadija Hesham

Mona Mansor

Gehad Abokamar

Dina Ibrahim

Agenda

1. Problem Definition
2. CNN Architecture
3. System Framework
4. Experiments and Results
5. Conclusion and Future Work
6. References

1. Problem Definition

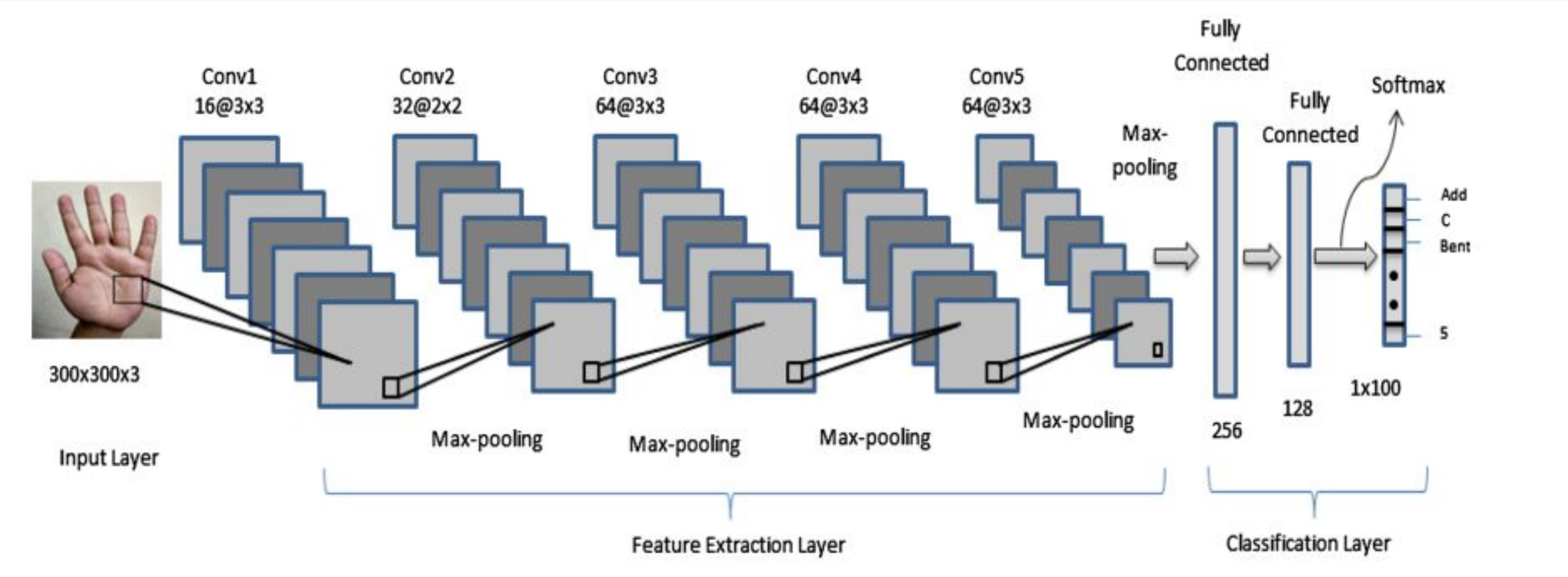
- The emerging need for human-computer interaction (HCI) without any input devices:
  - Speech and hearing impaired people
  - 5% (360 million approx.) of the total population of the world is suffering from either medium or severe hearing loss. [1]

- There are more than 300 different sign languages in use around the world (America, British, Chinese, French, …
  - Unfied sign language!

1. Problem Definition

- Applications:
  - Virtual Reality (VR)
  - Robot Control
  - Natural User Interfaces
- Static Hand gesture:
  - Position of hands and fingers in the space without any movement w.r.t the time
- Solution is a vision-based method:
  - Need a camera only!
  - Convolutional neural networks (CNN)

2. CNN Architecture



2. CNN Architecture

1-Convolutional layer

2-pooling layer

3-ReLU layer

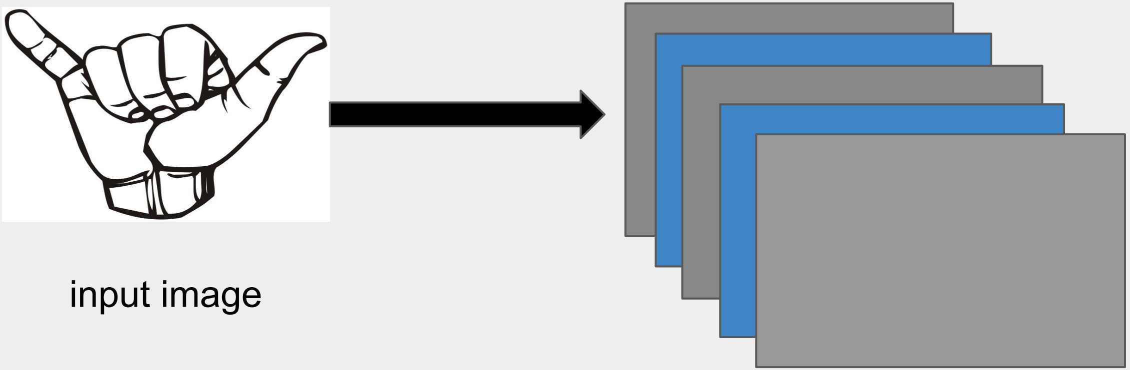
- 3.1-The sigmoid function
- 3.2-hyperbolic tangent

4-Fully connected layer/output layer(the classification layer)

2. CNN Architecture

1-Convolutional layer

$$\text{output} = \frac{W-F+2P}{S} + 1$$



input image

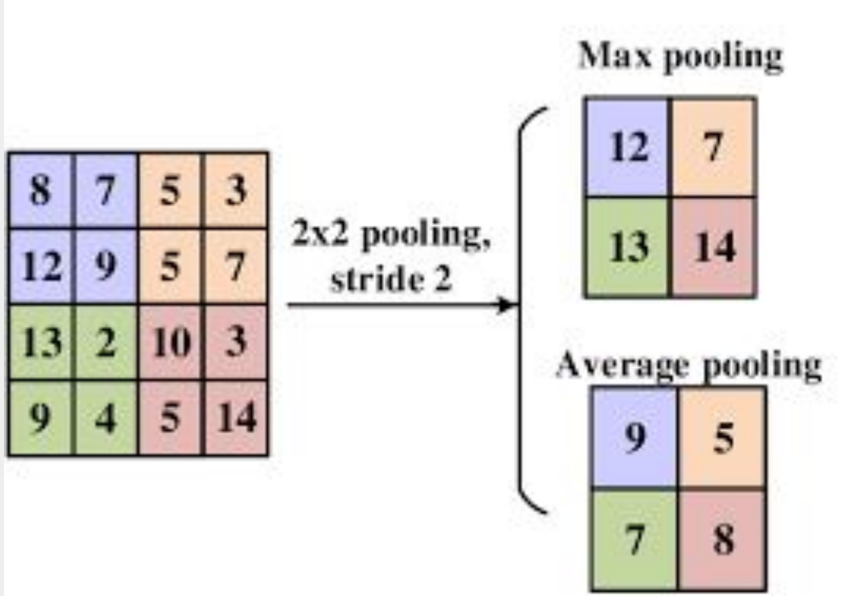
Conv 1

2. CNN Architecture

2-pooling layer

2.1-MAX Pooling

2.2 Average Pooling

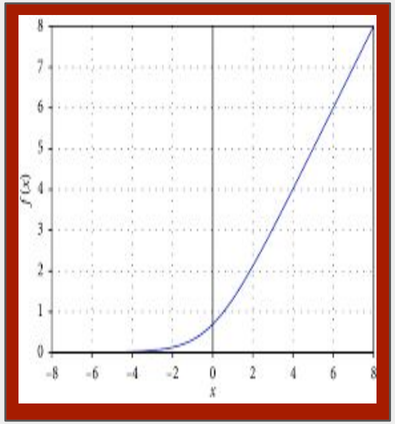


2. CNN Architecture

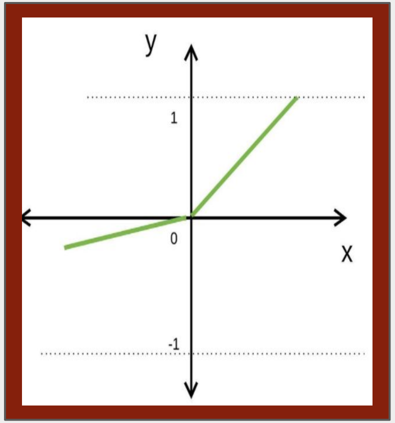
3-ReLU layer

<u>Some of the ReLU variants</u>

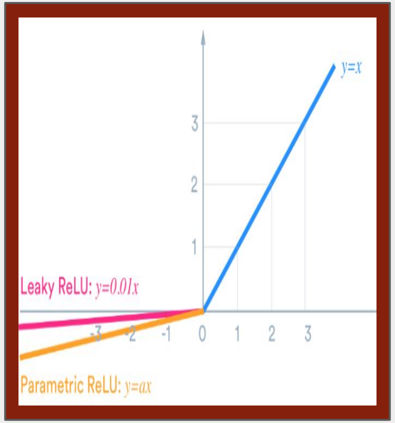
Softplus(SmoothReLU)



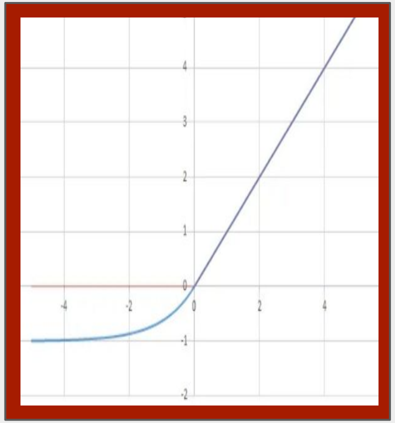
Noisy ReLU



Parametric ReLU



ExponentialReLU(ELU)

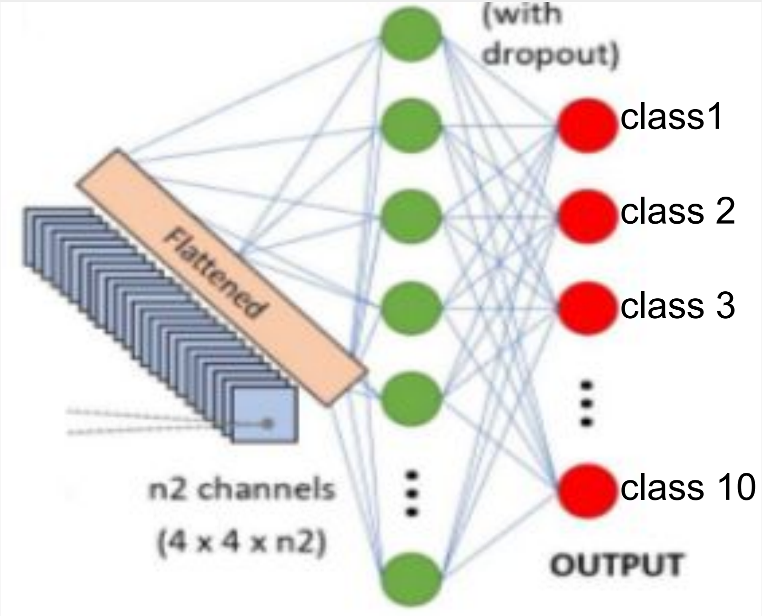


2. CNN Architecture

4-Fully connected layer/output layer

(the classification layer)

4.1 softmax



2. CNN Architecture

Overview of some models

1-LeNet-5 CNN_based model

2-Stacked Denoising Autoencoders network(SDAE)

3- Deep Belief Network (DBN)

4-Artificial neural network(ANN)

5-Custom CNN

6-VGG 19 NETWORK

3.1. Data

Data Acquisition

- Three channel RGB images.
- 35K image, 350 for each sign.
- 23 English alphabets.

Data Preprocessing

(Noise Removal)

- 128 x 128 Resizing.
- Normalization.

3.2. Modeling

Training

80% of the whole dataset.

Tools

- Tesla GPU
- 12 GB memory
- 64 GB RAM
- 100 GB SSD.

Testing

- 50 CNN, max 100 epochs.

- Loss function (cross-entropy)

- optimizers:

** Adaptive Moment Estimation (Adam).
** Adagrad.
** Adadelta.
** RMSprop
** Stochastic Gradient Descent (SGD).

4. Experiments and Results

Number of layers	Number of filters	Training accuracy (%)	Validation accuracy (%)	Number of epochs
8 (5 CL, 3FC)	16	10	5	100
5 (3 CL, 2FC)	16	42	26	100
4 (2 CL, 2FC)	16	99.17	98.80	20
4 (2 CL, 2FC)	32	98.82	98.53	20
4 (2 CL, 2FC)	64	99.05	98.76	20

Table 1 Experimental results with respect to parameters

Model	Training accuracy (%)	Training loss	Validation accuracy (%)	Validation loss	Optimizer
I	99.17	0.0280	98.80	0.0684	Adam
II	99.59	0.0378	98.27	0.1940	RMSProp
III	99.72	0.0126	98.56	0.0759	SGD

Table 2 Experimental results with respect to optimizer and colored images

4. Experiments and Results

Sign	Precision	Recall	F_1-score	Sign	Precision	Recall	F_1-score
A	1.00	0.96	0.98	Me	1.00	1.00	1.00
Afraid	0.97	0.97	0.97	Nose	0.98	1.00	0.99
B	1.00	1.00	1.00	Oath	1.00	1.00	1.00
Bent	0.97	1.00	0.99	Open	1.00	0.97	0.98
Coolie	0.97	0.94	0.96	P	1.00	0.97	0.98
Claw	1.00	1.00	1.00	Pray	1.00	1.00	1.00
D	0.79	0.97	0.87	Q	0.97	1.00	0.99
Doctor	0.98	1.00	0.99	S	0.95	1.00	0.97
Eight	0.96	0.90	0.93	Sick	1.00	1.00	1.00
Eye	1.00	1.00	1.00	Strong	0.97	1.00	0.98
Fever	0.95	1.00	0.97	T	0.99	1.00	0.99
Fist	0.97	0.98	0.97	Tongue	0.99	1.00	0.99
Gun	0.97	1.00	0.99	Trouble	1.00	0.95	0.97
H	1.00	1.00	1.00	U	1.00	0.99	0.99
Hand	0.97	1.00	0.98	V	1.00	1.00	1.00
I	1.00	1.00	1.00	West	1.00	0.93	0.96
Jain	0.99	1.00	0.99	Water	0.93	0.98	0.95

Table 3 Classification performance

4. Experiments and Results

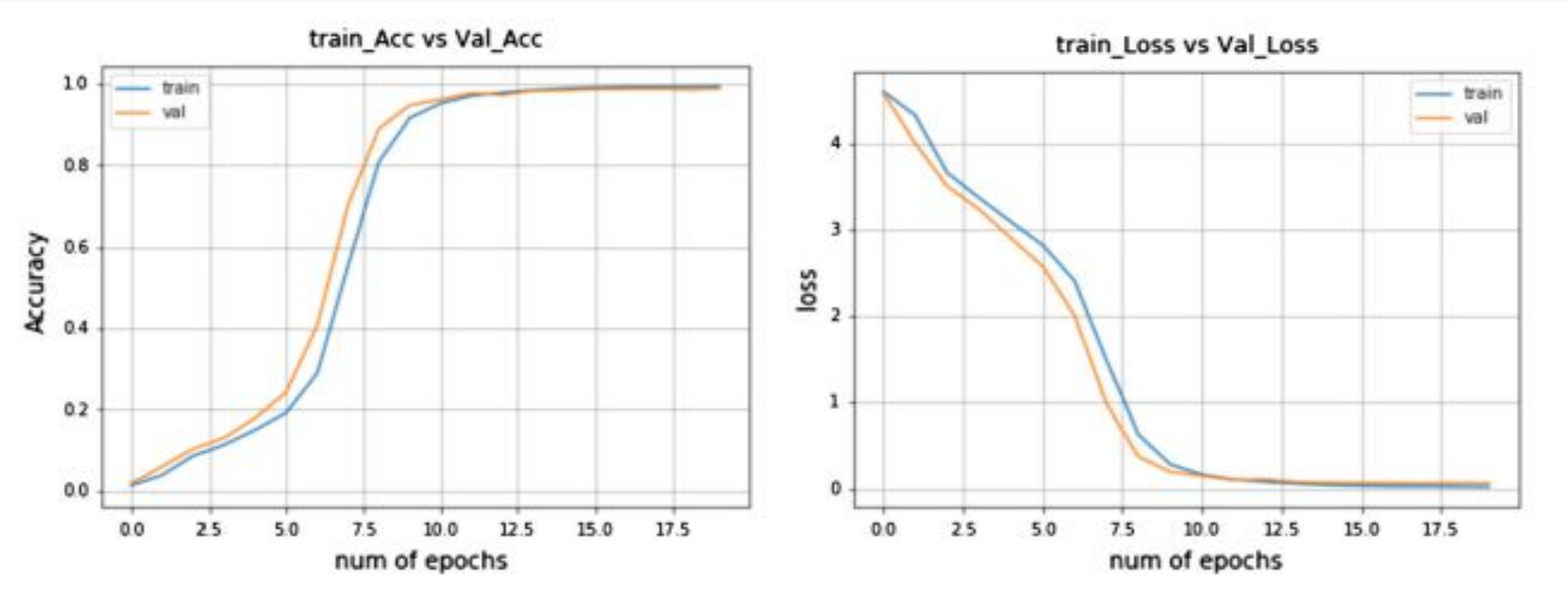


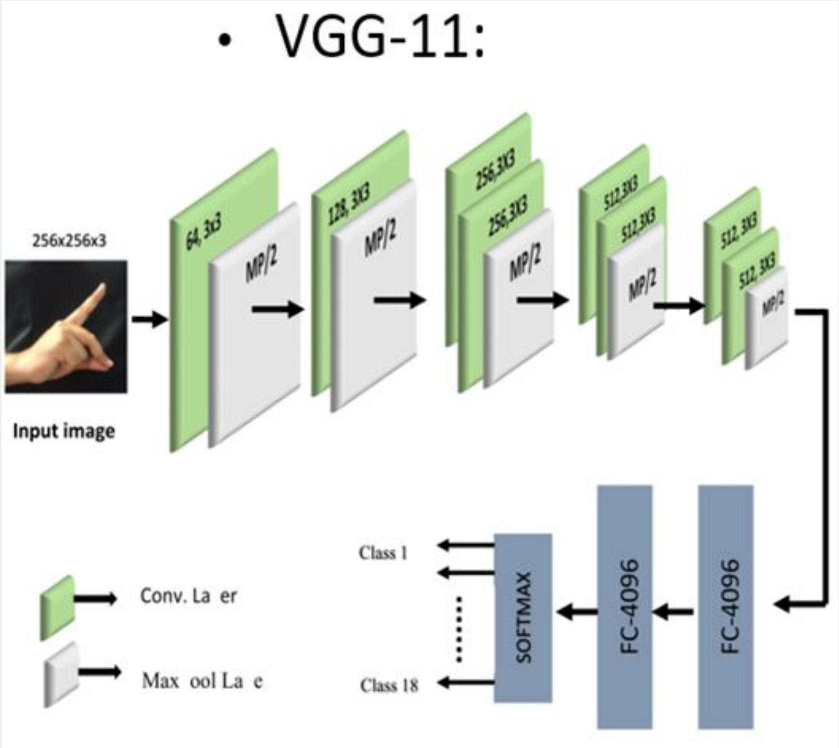
Fig 3.Accuracy and loss curves for training and validation datasets

5. Conclusion and Future Work

- Proposed system outperformed other existing systems even with less number of epochs. (highest training and validation accuracy of 99.17% and 98.80%)
- SGD optimizer outperformed other optimizers on grayscale images (training and validation accuracy of 99.90% and 98.70%)
- In addition to accuracy, try other efficiency measures to evaluate the robustness of the proposed system as we have larger number of hand gestures in real-world
  - Minimize computational complexity
  - Make classifier convergence faster
- Strengthen the CNN against overfitting by performing K-fold cross-validation
- Perform hyper-parameters tuning
  - Each dataset has its own best set of parameters

6. Other suggested solutions

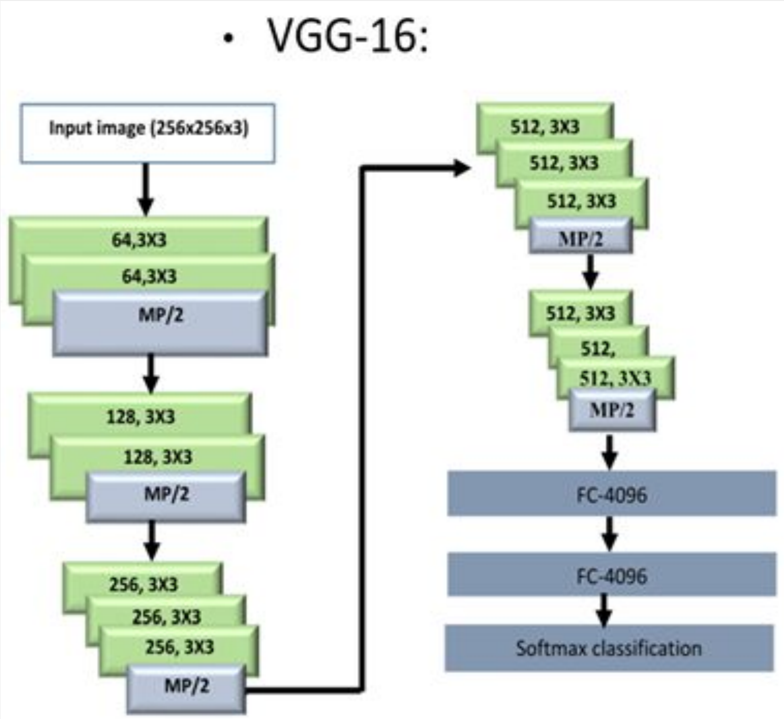
Accuracy : 98.54%



Reference : Vision-based hand gesture recognition using deep learning for theinterpretation of sign language

6. Other suggested solutions

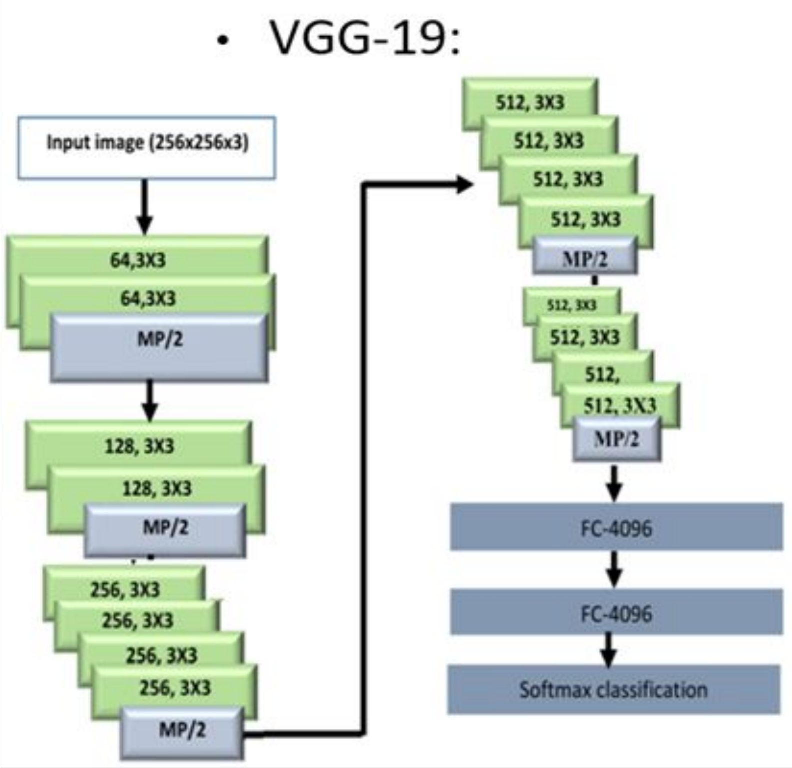
Accuracy : 97.12%



Reference : Vision-based hand gesture recognition using deep learning for theinterpretation of sign language

6. Other suggested solutions

Accuracy : 99.92%



Reference : Hand posture classification with convolutional neural networks on VGG-19 net Architecture

6.3. Myanmar Methodology

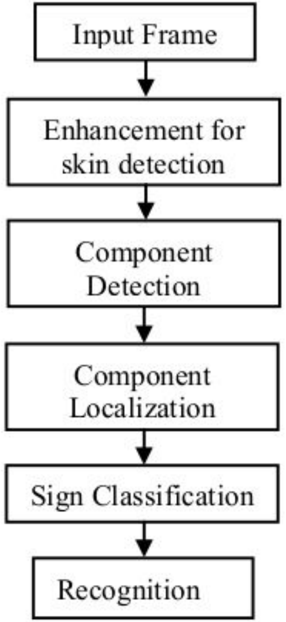


Figure 1. Block diagram of sign language recognition system

6.3. Myanmar Methodology

Group	Tested images	No. of correct image	No. of validate image	Percentage of correct	Percentage of error
R	2100	2010	90	95.71%	4.29%
L	600	581	19	96.83%	3.17%
C	3000	2896	104	96.53%	3.47%

Table1. Accuracy of Sign Classification and Recognition

ဟုတ်ကဲ့၊ စားပြီးပြီ။ (Yes.)	R12+ L4	
နေကောင်းလား။ (How are you?)	R1+R2	Doubl e group
မကောင်းဘူး။ (Bad.)	R4+C9	

Table2. Sample of Myanmar sign language groups

7. References

[1] https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[2] 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

[3] Sharma, S., & Singh, S. (2021). Vision-based hand gesture recognition using deep learning for the interpretation of sign language. *Expert Systems with Applications*, 182, 115657.

[4] S Amir et al (2020). Hand posture classification with convolutional neural networks on VGG-19 net Architecture. *IOP Conf. Ser.: Earth Environ.* Sci. 575 012186.

[5] S. M. Htet, B. Aye and M. M. Hein, "Myanmar Sign Language Classification using Deep Learning," *2020 International Conference on Advanced Information Technologies (ICAIT)*, 2020, pp. 200-205, doi: 10.1109/ICAIT51105.2020.9261775.


THANK YOU