

# TeleStroke System (TSS)

Rishabh Chandaliya, Praveen Joshi, and Haithem Afli

ADAPT Centre,  
Cork Institute of Technology, Cork, Ireland

**Abstract.** According to the World Health Organisation(WHO), cardiovascular diseases (CVDs) are the number 1 cause of death worldwide, claiming nearly 17.9 million lives yearly [1]. Eighty-five percent of all CVD deaths are from heart attacks and strokes. This rising crisis has received relatively small coverage up to date, given its massive economic development effect in countries. However, Early detection of stroke and treatment of it is essential for an excellent outcome. This research aims to discuss the challenge of providing early detection of stroke through drooping mouth detection on Client-Server based architecture, a Tele-Stroke System (TSS), which is helpful for initial treatments. In this paper, the input data is from the Kaggle and YouTube Facial Palsy (YFP) database. We can find that simple Machine learning gives better accuracy than other deep learning models from the experimental results. It means that we can effectively assist doctors in early detection and diagnosing of stroke.

**Keywords:** Stroke Detection · Data Augmentation · face detection · facial palsy · droopy mouth detection · Machine Learning.

## 1 Introduction

Stroke is commonly described as a clinical syndrome of presumed vascular origin, characterized by rapid signs of local or global cerebral dysfunction that usually lasts for more than 24 hours or even death [18]. A stroke is said to occur when a blood vessel carrying nutrients and oxygen into the brain is either blocked, bursts, or ruptured by a clot. This results in decreased oxygen and blood supply in the brain, causing behavioral problems, memory loss, thinking habits, impairment, permanent brain damage, and even death [2].

Generally, there are three types of stroke. The first type of stroke is called Ischemic stroke. It occurs when there is a blockage in arteries either by gradual plaque build-up, blood clots, and other fat deposits. The second form is the hemorrhagic stroke that happens when a blood vessel located in the brain splits and induces blood to leak into the brain. This form of stroke is responsible for 13 percent of all strokes, and it accounts for over 30 percent of all stroke-related deaths. Lastly, an Embolic stroke occurs when a clot breaks off from the artery wall and becomes an embolus that can travel further down the bloodstream, resulting in the smaller artery's blockage. Embolic typically originates from the heart, where uncommon diseases may cause clot formation.

Depending on the latest National Stroke Association statistics[3], stroke is now the fifth leading cause of death in the United States alone and is a significant cause of adult disability. Every year, 0.8 million people in the U.S. are believed to have a stroke, and 0.13 million of them died from a stroke, meaning, on average, one American dies from stroke every 4 minutes. The lack of understanding of stroke signs is also one of the contributing factors for the unprecedented increase in death associated with the stroke.

## 2 Related Work

When you deal with stroke, it's always crucial to know what you want to identify. It is safer to provide a standardized program that addresses specific common stroke symptoms. One of the signs of stroke is facial paralysis, meaning, on one side of the face, there is facial sagging, which is known as face drooping. The droopy mouth is not only a symptom of stroke but also a sign of a brain and nerve disorder known as Bell's palsy. Bell's palsy, in short, is muscle weakness or paralysis on one side of the neck. As a result, facial nerve damage that controls the muscles on one side of the face causes that other side of the face to drop[4].

A Vision-based interface is provided where the user's hand, foot, or head movement are identified and monitored. The framework is designed for child-to-computer interaction. They base the device assessment on the concept of HAAT (Human Activity Assistive Technology) [5].

It projects the system to be less intrusive and cheaper than other solutions, in which they built up a video-based prediction process. This function: first, the movements of different parts of the body are isolated, then the motion characteristics are extracted and used to identify persons as stable or impaired. The results show that visually collected motion data allows the Identification of Bell palsy and Pseudoperipheral palsy as the state-of-the-art detection using data from electromagnetic sensors [6].

" Camera Mouse" was introduced in [7] to provide computer access for people with severe disabilities. With a video camera, the device monitors the user's gestures and converts them into the mouse pointer, which moves on the screen.

They built a visual monitoring system in [8] that passively observes moving objects in a site and learns from these observations patterns of behavior. The system is based on motion tracking, camera synchronization, recognition of operation, and events. The method is useful for sequence classification as well as for individual site operation instances.

In [9], they have converted a video into a sequence of frames used by correlating structures to predict human behavior. The machine examines the faces and identifies them to test the expression ROI (Region of Interest). It also utilizes Viola-Jones patterns, and the AdaBoost process to filter colors to identify the ears, noses, eyes, mouth, or upper body of people. These sections have higher entropy to detect emotions.

A method was proposed that recognizes three facial expressions in [10]; they use the geometric feature approach for extraction of features; the system uses the

neural network algorithm Multilayer Perceptron (MLP) with backpropagation for classification. The facial expressions to be identified are optimistic, happy, and surprised. The overall identification rate is 93.33 percent.

They provide a system for recognizing facial emotions in [11]; it is based on ASM (Active Shape Model) for facial feature extraction and Radial Basis Function Network (RBFN) to evaluate the symmetry of the face in different behavior patterns with accuracy 90.73

In [12], they developed a computer vision system that uses hidden Markov models (HMMs) to automatically identify individual action units or combinations of action units. Researchers used three methods to extract information on facial expression:

- Tracking point
- Complex flow monitoring with principal component analysis (PCA)
- Identification of high gradient components.

They present a system for face recognition and face emotion detection in [13], using the Open Source Computer Vision (OpenCV) library and python machine learning. Machine learning algorithms are used to recognize and classify different facial expressions and body movements. (OPPOSE)

The research focuses on image processing to obtain facial features/landmarks, mainly mouth corners, determine the droopy mouth threshold value and deployment using Android Studio. Consequently, the mobile application will address certain user groups such as neurosurgeons and emergency medical services. However, the mobile application could also be useful for patients, potential patients with strokes, and generally any user who is health-conscious anywhere and anytime[14].

## 3 Material

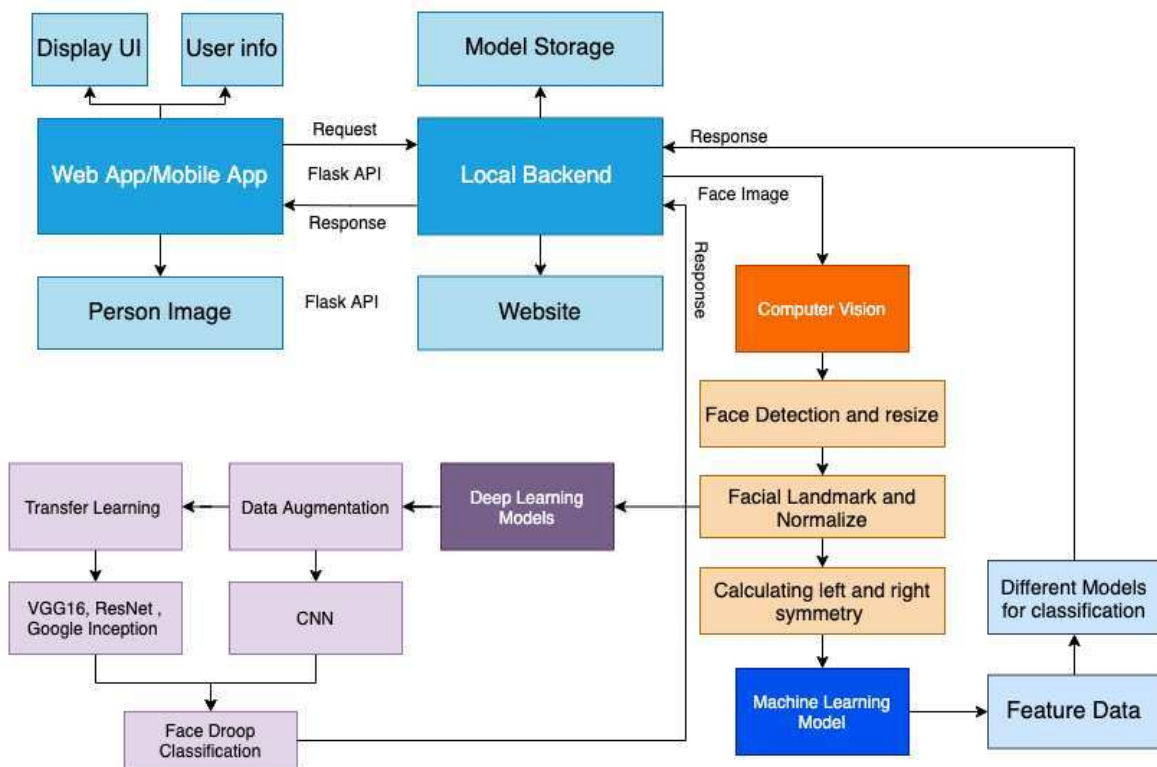
### 3.1 DATASET

- **Stroke Faces on Kaggle[15]**: This dataset consists of 1000 droopy faces curated from Google by Kaitav Mehta.
- **YouTube-Facial-Palsy-Database[16]**: The dataset in this paper contains 32 video clips collected from YouTube of 22 patients with facial paralysis. It is the first public database made available for the visual inspection study of facial palsy symptoms. Gee-Sern Jison Hsu, Jiunn-Horng Kang, Wen-Fong Huang[16] convert the videos into image sequences and have the images labeled by clinicians with facial palsy. As the number of patients is small, we have removed the patient's multiple pictures from the video.
- **Yale Face Database B[17]**: The collection comprises 5760 single-source light photographs of 10 subjects seen under 576 viewing conditions each. The subjects are healthy and do not have any symptoms of a stroke. A single image of subjects from the dataset is used for regular face input in CNN feed.

Number	Stroke	No Stroke
Training Images	900	900
Testing Images	100	100
<b>Total</b>	1000	1000

## 4 Method

This segment addresses the proposed TeleStroke system for Eye area and the mouth region. TeleStroke consists of client-server stages: the client-side relates to video pre-processing using face detection and landmark position to locate each sequence frame's faces. In this phase, the Integrated Deep Model is used. The face images detected are then cropped to the face, and the number of frames per sequence is normalized to a fixed length. The server side consists of various models, such as the ML model, FCNN, CNN, VGG16(Transfer Learning), one for each of the face analysis tasks. The proposed structure for Telestroke is given in figure 1.

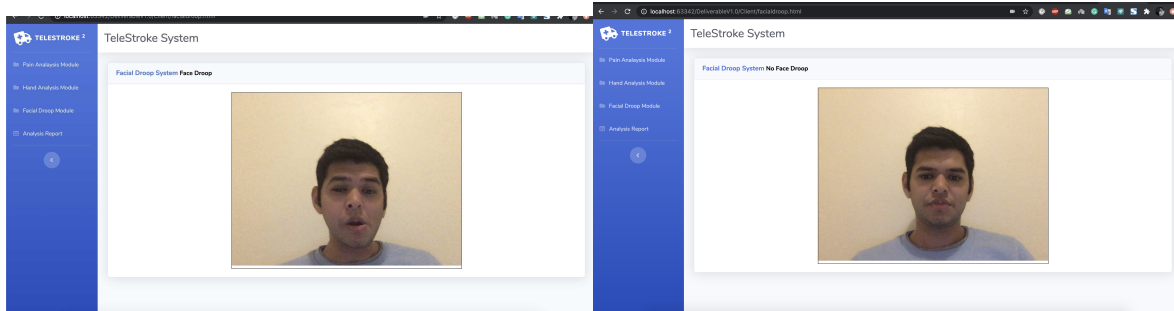


**Fig. 1.** Deep Client-Server Architecture

### 4.1 Front End

The front end is a web application comprised of HTML, CSS, JavaScript, AJAX, which ask the user's permission to grant for camera access that takes live stream

image of the user i.e., a WebSocket connection is established. From this stream, a picture is taken out, and the build model is applied. Ajax call is made to get the response and request faster.



**Fig. 2.** Face - Droop Image

**Fig. 3.** No Face - Droop Image

## 5 Experimental Evaluation

This section provides a detailed experimental evaluation of the Stroke recognition system using the TeleStroke Framework. All experiments are conducted on Google Colab with GPU using the Tensorflow 2.0.

The dataset is combined and viewed as one group of droopy as 0 and not droopy faces as 1. This dataset was transferred to different steps and tested on various algorithmic models.

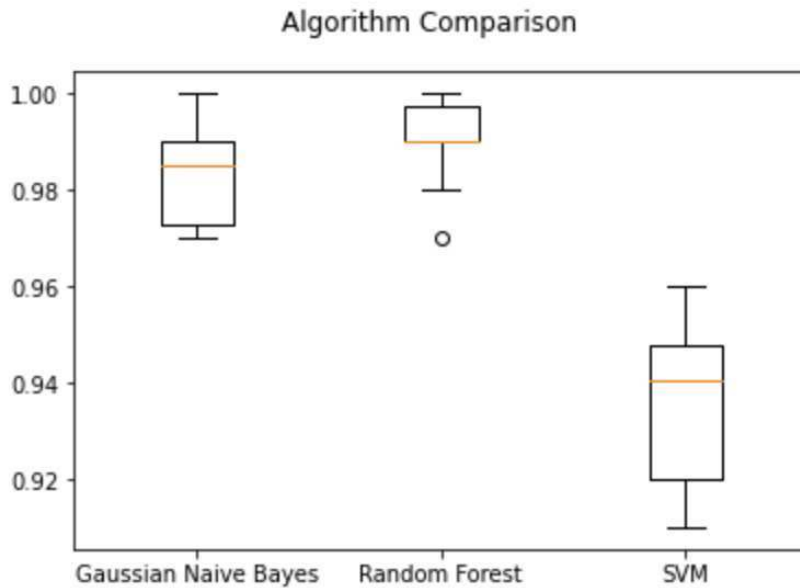
The evaluation methodologies are as follows:

### 5.1 Simple ML model

In this part, the Area ratio of the eye and Slope calculation of mouth plus eye and mouth coordinates are saved in CSV file, and the droopy faces are labeled as 0, and Normal Subjects are marked as 1 which will be used for classification.

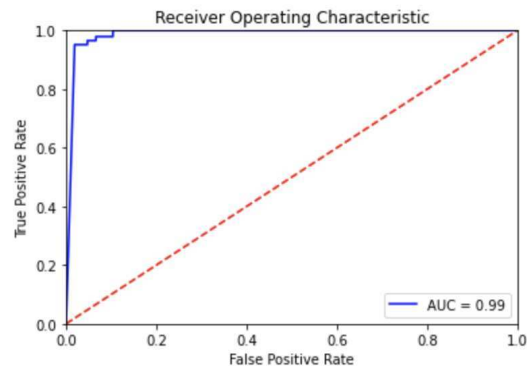
The Fig 4. shows the 10-fold cross-validation technique is performed to evaluate each algorithm on training data, which is critically designed with the same random seed to ensure that the equal splits are presented to the training data. Each algorithm is evaluated precisely the same way.

Fig 5 and 6 is the precision, recall, f1 score, and ROC Curve for the Naive Bayes as it was only tested with the testing data. It would indicate from these results of 10 cross-fold validation both Naive Bayes and Random Forest are perhaps worthy of further study on this topic.



**Fig. 4.** Algorithmic Comparison

	precision	recall	f1-score	support
0	0.95	0.93	0.94	109
1	0.94	0.96	0.95	142
accuracy			0.95	251
macro avg	0.95	0.95	0.95	251
weighted avg	0.95	0.95	0.95	251



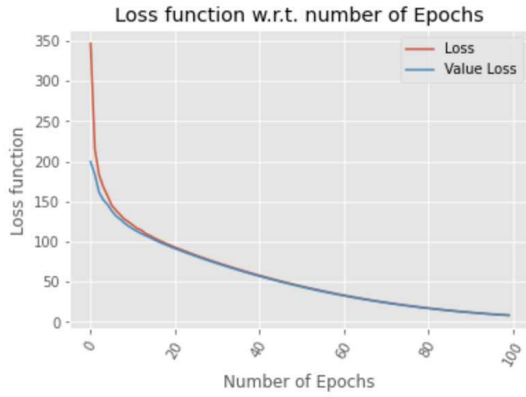
**Fig. 5.** Precision, Recall, F1-Score

**Fig. 6.** ROC Curve

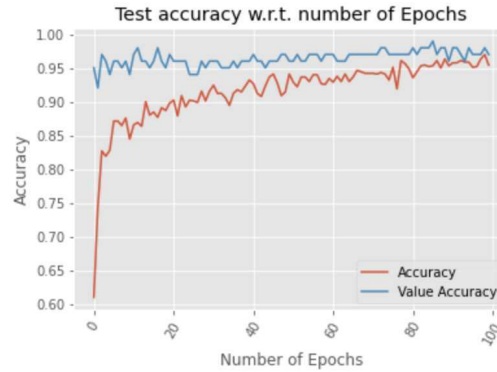
## 5.2 Fully Connected Neural Network.

The first layer has 600 neurons, the second layer has 400 neurons, and the third layer comprises of 200 neurons as shown in Fig 4.7 The lower the loss, the better a model will be (unless the model is overfitted to the training data). The loss is based on training and validation, and its definition for these two sets is how well the model is doing. Contrary to accuracy, the loss is not a number. It is a summation of the errors in training or validation sets that were made for each example.

There are also other subtleties when raising the value of the loss. For example, we may run into the problem of over-fitting, in which the model "memorizes" the patterns of training and becomes ineffective for the test set. We use Dropout for that reason.



**Fig. 7.** Loss Function

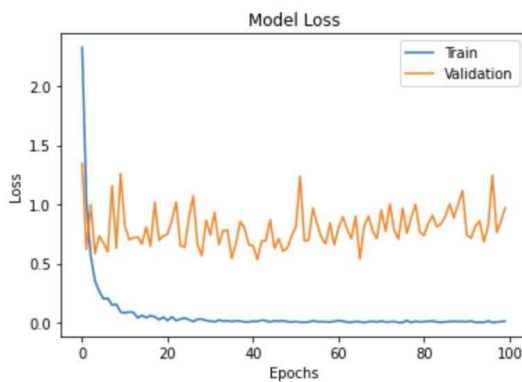


**Fig. 8.** Accuracy

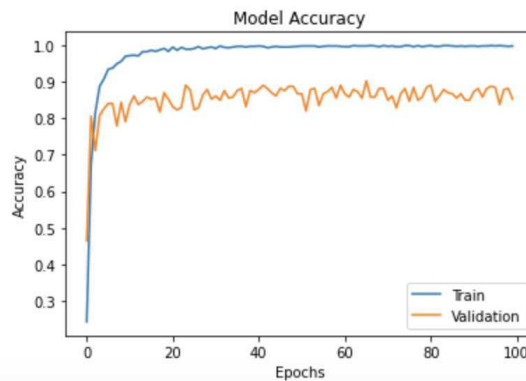
Name	Epoch	Accuracy	Loss
FCNN	100	0.95-0.97	0-2

### 5.3 CNN and VGG16 model

The amount of "wiggle" in the loss is proportional to the size of the batch. The wiggle should be reasonably large when the batch size is 1. If the batch size is the maximum data-set, the wiggle should be small, as each gradient update will monotonically increase the loss function (unless the learning rate is set too high).



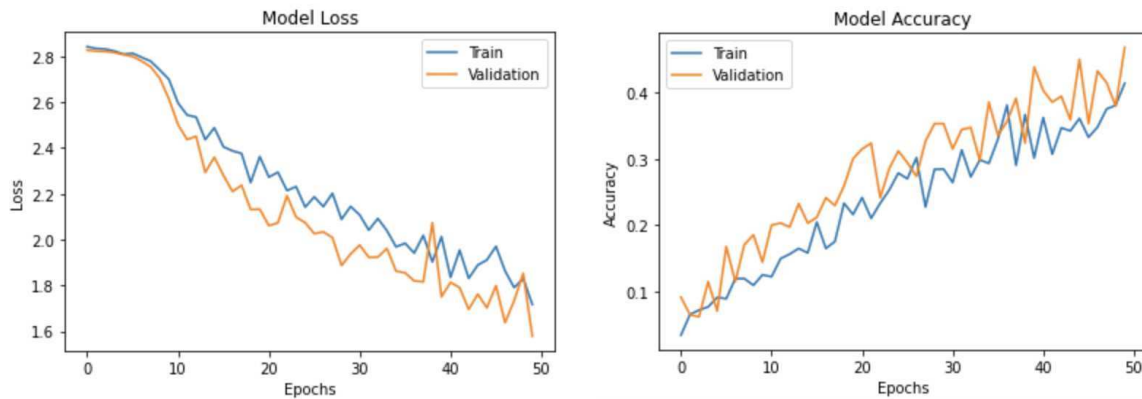
**Fig. 9.** Loss without augmentation



**Fig. 10.** Training loss without augmentation

The difference between the accuracy of the training and evaluation indicates how much overfitting. In Fig 7 and 8, we can observe that too much wiggle is formed, and the model starts to overfit at 0-20 range. The data is not augmented.

The yellow validation curve shows very low validation accuracy relative to the training's accuracy, suggesting heavy overfitting (note that validation accuracy can also begin to go down after some point). Compared to the above Fig 7 and 8



**Fig. 11.** Loss with augmentation

**Fig. 12.** Training loss with augmentation

Fig 9 and 10 is better and the wiggling is less and there is no sign of over fitting.

## 6 Conclusion

The proposed Web application requires incorporating four critical processes: image acquisition, key points extraction using OpenCV, mathematical measurement, and droopy mouth detection. This paper uses a Simple ML, Fully Connected NN, and CNN architecture with ResNet backbone to present a fully end-to-end framework named Telestroke System for droopy mouth detection. We tested the design using a combination of 3 datasets to get an F1 score of 94 percent, respectively, for the droopy mouth.

## 7 Discussion and Future Work

In this paper, we took on a novel and challenging task of assessing the different models with our proposed model. Although our model performed very well on the selected data sets, there are many areas where it is possible to pursue this research. Firstly, it is possible to achieve hyperparameter tuning and optimization by adding some complex feature vector representations; there is also the potential to investigate more complex backbone and deeper networks, which are limited in this work due to computational overhead 3D CNNs. It opens up a research opportunity for simplifying droopy mouth detection & recognition in Mobile Application.



The current dataset should be expanded to a more consistent one study by including a more substantial number of face photos labeled with demographic details, including age, gender, and ethnicity. As the data collection expands, additional knowledge will become more accurate, and we will consider studying the impact of original populations on paralysis disease and its ranking. Besides, the performance comparison of different pre-trained CNN models can be improved by using more precise facial landmark detection techniques and enabling the recognition of a droopy mouth from tilted or rotated images. Besides, the proposed model could be enhanced to achieve faster identification of droopy mouth and better recognition performance. Will be seen as a potential problem for research.

## 8 Acknowledgments

This research was conducted with the financial support of ADAPT Core (Platform & Spokes) under Grant Agreement No. 13/RC/2106 and at the ADAPT SFI Research Centre at Cork Institute Of Technology. Science Foundation Ireland funds the ADAPT SFI Centre for Digital Media Technology through the SFI Research Centres Programme. It is co-funded under the European Regional Development Fund (ERDF) through Grant 13/RC/2106.

## References

1. Stroke: a global response is needed <https://www.who.int/bulletin/volumes/94/9/16-181636.pdf>
2. MayField Clinic. 2013. What Is Stroke? [http://www.mayfieldclinic.com/pe-stroke.htm#.Va0rm\\_mqqko](http://www.mayfieldclinic.com/pe-stroke.htm#.Va0rm_mqqko)
3. National Stroke Association. 2016. Signs and Symptoms of Stroke
4. The Nature of Bell's Palsy. (Laryngoscope. 1949;59:228-235), R. Gacek, "Hilger
5. Vision based interface: an alternative tool for children with cerebral palsy, Magdalena Gonzalez, Debora Mulet, Elisa Perez, Carlos Soria, and Vicente Mut.
6. Video-based early cerebral palsy prediction using motion segmentation, Hodjat Rahmati, Ole Morten Aamo, Øyvind Stavdahl, Ralf Dragon, and Lars Adde
7. The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities, Margrit Betke, James Gips, and Peter Fleming
8. Learning patterns of activity using real-time tracking, Chris Stauffer and W. Eric L. Grimson
9. Human behavior prediction using facial expression analysis, Subarna Shakya, Suman Sharma, and Abinash Basnet
10. Real time facial expression recognition using realsense camera and ann, Jayashree V Patil and Preeti Bailke
11. Facial emotional expressions recognition based on active shape model and radial basis function network, Endang Setyati, Yoyon K Suprpto, and Mauridhi Hery Purnomo
12. Automated facial expression recognition based on face action units, James J Lien, Takeo Kanade, Jeffrey F Cohn, and Ching-Chung Li

13. A robust method for face recognition and face emotion detection system using support vector machines, KM Rajesh and M Naveenkumar
14. Mobile Health Awareness in Pre-Detection of Mild Stroke Symptoms, O.-M. Foong, J.-M. Yong, S. Sulaiman and D. Rambli
15. Dataset for Facial Palsy Kaitav Mehta =<https://www.kaggle.com/kaitavmehta/facial-droop-and-facial-paralysis-image>,
16. Deep Hierarchical Network With Line Segment Learning for Quantitative Analysis of Facial Palsy
17. Acquiring Linear Sub spaces for Face Recognition under Variable Lighting
18. World Health Organisation. 2014. Stroke, Cerebrovascular, [http://www.who.int/topics/cerebrovascular\\_accident/en](http://www.who.int/topics/cerebrovascular_accident/en)