DEEP LEARNING APPROACH TO DETECT FACIAL ASYMMETRY FOR PARALYSIS DETECTION

SAMUEL SUSAN VEERAVALLI[1] , DR. PRAJNA BODAPATI[2]

[1]Research Scholar, Department of CS & SE, Andhra University College of Engineering, Andhra University, India.
[2]Professor, Department Of CS&SE, Andhra University College of Engineering, Andhra University, India.
Email: [1]vssusan.rs@andhrauniversity.edu.in [2] prof.bprajna@andhrauniversity.edu.in

ABSTRACT

Facial paralysis is the inability to contract the muscle nerve on one (or both) of the sides of face, which directly results in deforming or a droopy face. This is considered as one of the diseases that is increasing its rate of infection in recent times. There are many causes for Facial paralysis and is also a common symptom in people who have suffered a brain stroke. Although, there are vast applications with the help of technology is being used in different areas of medical diagnosis, there are very few notable models for detection of facial paralysis available presently. So, this work proposes an efficient model with the help of deep learning models which were trained and known to be best for image analysis. In this, the input is a dataset consisting of images that are normal faces and paralyzed, different models are implemented of those to classify them. The algorithms which resulted in obtaining the best accuracy are DenseNet201, VGG19 and InceptionResNetV2, with an accuracy of 99.84%, 99.51%and 99.19% respectively.

Keywords:- *Facial paralysis, image classification, DenseNet201, VGG19 and Inception Resnet.*

I. INTRODUCTION

In Today's World, Statistics have proven that there has been an increasing rate of diseases in humans, in which facial paralysis is one disease of concern though relatively rare, having a population-wide yearly incidence of roughly 30 per 100,000 people [1]. Facial palsy, also known as facial nerve paralysis, is a condition in which the facial muscles weaken as a result of injury to the facial nerve that may be temporary or permanent.

If the facial nerve is damaged or absent, the facial muscles do not receive the signals they require to function properly. This can result in paralysis of the affected area of the face, impairing movement of the eyes, mouth, and other areas. Facial paralysis can cause facial asymmetry, resulting in a loss of structured appearance and function, which can be detrimental to the patient's psychology and social life.

Facial paralysis comes in varying degrees. The bottom half of the face is sometimes affected, the entire face is sometimes affected, and both sides of the face are sometimes affected. Since each side of the face has a different facial nerve, damage to one will only impact that side of the face, and vice versa, when it comes to the effects of facial palsy. Each nerve has a brain origin before emerging from the face and splitting into five distinct branches in front of the ears. The muscles required for facial expressions are provided by these branches. The facial nerve regulates taste, salivation, and tears in a few different ways.

Facial Paralysis has various types and can be caused in both adults and children at equal degree, Although Bell's Palsy[2] is seen to be the most frequently seen type which is an idiopathic condition, with a rate of 54.9% in adults and 66.2% in children, then comes infection based paralysis that is usually caused by a virus known as Ramsay Hunt Syndrome was seen at 26.8% in adults and 14.6% in children, traumas like after effects of strokes, surgeries, child birth traumas, accidents etc. were seen to be the cause of paralysis at 5.9% in adults and 13.4% in children which includes 3.4% caused due to birth separately, followed by iatrogenic causes in adults at 2% and last but not least caused due to tumors at 1.8% in adults and leukemia in children at 1.3%.Also in few cases, the facial nerve may be damaged due to external injuries caused during accidents.

In this work, using deep learning models for image classification to detect asymmetry in faces, earlier works used different approaches and dataset. The objective is to detect facial palsy. Hence a CNN based models, DenseNet201, VGG19 and InceptionResNetV2 are used for training a model in detection of facial palsy by using the images containing faces of people who are affected with paralysis on face. Superior accuracy versus earlier approaches. This model only works for detection, it is unable to assess the degree of deformation in faces or evaluate the recovery stage.

2. LITERATURE REVIEW

Gemma S.Parra-Dominguez et al. developed a facial paralysis detection system [3] that is a conventional system that begins with extracting facial landmarks, due to prior knowledge of not having proper accuracy in doing so they took benefit of a guarin-developed 68 key point shape predictor, which was used in the MEEshape predictor that was seen to have better prediction performance, the process of extraction involved converting the inputs of color-scale to gray scale, Resizing using scale factor sf = W/nW where nW = 200 and nH = sf in which W and H indicate the width and height, later face detection on the resized image is done, rescale using sf and lastly, predict and store the data that is extracted for data processing in the future, MEE shape predictor does training to detect 68 points, but only 51 were used in this model, this model also employs to do tilt correction using getRotationMatrix2D as accuracy in finding symmetry of the face can be effected by the tilt of the head if the position is tilted, t by considering the first and the last key points on the jaw structure of the face the tilt correction can be done.

Rishabh Chandaliya et al proposed a TeleStroke System which uses Straightforward Machine learning, Fully Connected NN and CNN architecture with ResNet as its backbone [4] where detection of stroke at an early stage and treatment is said to be essential for an extraordinary and accurate outcome, it is done through drooping mouth detection on Client-Server based architecture which is the TeleStroke System that can be said to be useful during early stages of treatment, the datasets for Stroke faces are taken from Kaggle that consist of 1000 droopy faces, Youtube-Facial-Palsy-Database, that has 32 video clips taken from YouTube that has inputs from 22 patients that suffer from facial palsy and Yale Face Database, that comprises of 5760 single-source light photographs of 10 patients seen under proper viewing conditions. This system mostly focuses on Paralysis in the eye area and the mouth region by using 10-fold cross validation technique, here f1 score and ROC Curve for the Naive Bayes are tested, TeleStroke consists of stages in which the first phase involves key point/landmark extraction, related to client side where video processing using. Face-detection and landmark position is done, it uses an Integrated Deep Model and later the face Images detected are resized and number of frames per sequence are normalized.

Gee-Sern Jison Hsu et al. Proposed a Hierarchical Detection Network (HDN) [5] which is said to be inaugural facial palsy detection approach in deep learning, the proposed method is made up of a three component network, first network is used to detect faces, second network detects the landmarks on faces used to detect facial palsy, and the third network detects the local regions effected by facial palsy, the first and third component networks use Darknet frameworks called FaceNet and PalsyNet denoted by Net f and Net p respectively , but it is used with fewer convolutional layers for reduced computation speed, the second component network uses the 3D face alignment network for finding landmarks denoted by Net m,outputs are based on YOLO-9000(state-of-the-art-real-time object detectors) proposed by Redmon and Farhadi and are trained using the dataset WIDER FACE, which contains 393,703 labelled faces in which 32,203 of images have a wide range of pose, lighting, expression, scale, and occlusion and results in an AP of 99.25% on AFW benchmark. Net m after face detection is employed using the state of art technique: Facial Alignment Network (FAN) that integrates the techniques Hour-Glass network that is made up of several stacked hourglass modules that allow for bottom-up, top-down inference in iteration and is based on simple nearest neighbor up sampling, which is a latest CNN for human pose estimation and a revolutionary residual block. This method examined 32 videos of 21 patients who suffered from facial palsy from YouTube, labelled all of the statistics via way of means of 3 hospitalists, and could make this database to be had to the studies community. Accuracy of this model is compared with and without CK+ and is seen to have 89% at recall 87% and 93% at recall 88% respectively in terms of precision.

Jocelyn Barbosa et al. proposed a method where the front view facial images of the patients that are significantly illuminated were used, four different positions of the face were examined first, starting off with the resting position, followed by the positioning of eyebrow and nose angling for different expressions, like smiling[6].In this method raw images are retrieved from a database where dimension alignment and resizing of images is done, later preprocessed to enhance contrast and get rid of any noisy images, facial future deformation is processed by high-pass gradient or Gabor wavelength-based filters by extracting salient facial points using Histogram of Oriented Gradients (HOG) integrated with a linear classifier, an image pyramid, and detection scheme using sliding window feature partitions the image into cells and gradient in histogram is calculated based on which they are discretised into angular bins, here the objective is to locate the 68 key points on the boundaries of facial features. In their work, they have concentrated more in obtaining a geometric based symmetry for the points that are generated for each image: mouth angle (MA), infra orbital (IO), upper eyelids (UE), supra-orbital (SO), nose tip (NT) and nostrils using dlib library, also uses ensemble of regression tree and shape invariant split tests which examines the intensities between two pixels positioned at each node in regression tree. Overall, this method uses 440 facial images from 110 subjects where 70% of dataset is used as training set and 30% is used as test set and states that ERT model has appealing quality and reduced error and this method significantly reduces time complexity of extracting features, without compromising on accuracy for real applications.

Chaoqun Jiang et al proposed an automatic facial paralysis assessment that equips computational image analysis where the flow of blood in the face of people suffering from facial paralysis[7] is taken into account and measured by laser speckle technique where a contrast is generated for both RGB images and blood flow images, and later uses and is a better approach for segmentation, where the patients face is divided into regions of concern to study the characteristics of how blood flow is distributed. The second phase in this study is camera estimation, which determines the camera model at where it is positioned using the set of 2D coordinates from face landmarks pi R2, I 0, 1, ..., 67 and their known correlations from the Surrey face model. The gold standard algorithm can be used to estimate the matrix of an affine camera model. The points generated from the D model that correspond to the 2D face landmark points pi are represented in homogeneous coordinates xi R3 and Xi R4, respectively. It is possible to get the estimated camera matrix C R34 using the gold standard algorithm. This method is validated on 80 FP patients and has received an accuracy of 97.14% on experimentation.

3. METHODOLOGY

3.1 Dataset

In this work, the data that needed to be given as input are the images which are definitive and uncomplicated. Considering that features for training and validating the model using deep learning, two datasets are used, "Facial Droop and facial paralysis" [15] dataset from Kaggle for facial paralysis images. Secondly the UTK Face Cropped dataset [16] which contains normal face images required for comparison. Among those, 800 images of stroke faces and 1000 images containing normal face are considered for data labeling. To obtain an 80-20 ratio for the validation, 200 images from both the datasets are labeled.

3.2. Overview Of Method

There are several methods proposed for detection of facial paralysis using deep learning methods, with image analysis [7][8][9][10][11]. In this work, the first step is collection of images from different datasets containing facial images with and without facial palsy. Then the basic data processing step is done on those images, i.e., all the images are scaled to the size of 224 x 224 pixel, which is required for the model. From the available images 80% of them are used for training the model and the rest are used for testing and validation of the model. The performance of different models is then compared with the results that are obtained after validation. In this work, to implement the model python frameworks Tensorflow and Keras were used. The implementation of the following models improved the detection rate.

For the detection of facial palsy different trained models are used, they are VGG19, DenseNet201 and InceptionResNetV2. These models have given the high accuracy for the existing classification models where images are given as input, figure 01. Hence these models are trained with the dataset containing droopy facial images and the normal facial images.

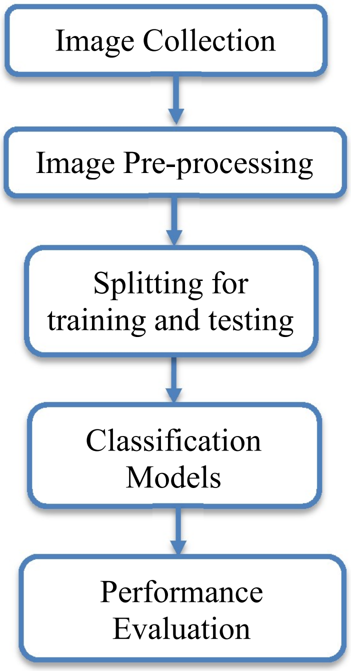


Figure 01: Methodology

3.2.1. DenseNet201 Algorithm

There are many research works that used the basic CNN models [12], where the layers are connected step by step as in formula (1). DenseNet201, is also one of the CNN based model which has 201 layers, this is well known model among all the newly proposed for its amazing performance on competitive object recognition, despite having fewer number of parameters. This model is earlier used in classification of COVID-19 patients with images as input [12]. In general, as the network goes deep it is difficult for classification as the gradients will be diminishing and will be vanished at some point. This problem is earlier discussed in ResNet model, the concept of ResNet is to ignore a couple of layers and create shortcut connections, as shown in figure 02, which can be effective. For example, if the input is x_{i-1} and assuming the output is after two layers $H(x_{i-1})$ is added to input layer x_{i-1}. The Equation (2) give the output for the ith layer. Such a shortcut branch is used to overcome the vanishing gradients in the case of deep networks in an inception module that is improved. To enhance the training process smoothly in deep networks that are based on gradient, highway networks were designed which allow the flow of information from previous layers to next layers without loss.

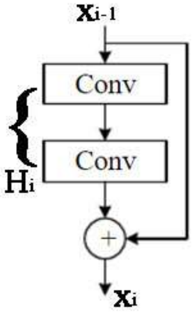


Figure 02: ResNet Block

DenseNet model improves the gradient propagation by creating short paths from early layers with the later layers. For example, if there are ten layers in the network, the tenth layer gets the input from all the previous nine layers. So, if there are N number of layers, there will be N*(N+1)/2 connections, Figure (03). While using deep and wide architectures to improve deep CNN performance can be beneficial, DenseNet is the updated model of the previous model, instead of adding the previous layers, in this the layers are concatenated in an orderly manner, which can be given as an equation given in (3).

$$x_i = H_i(x_{i-1}) \quad (1)$$

$$x_i = H_i(x_{i-1}) + x_{i-1} \quad (2)$$

$$x_i = H_i([x_0 + x_1 + x_2 + \ldots + x_{i-1}]), \quad (3)$$

where i is the index if the layer, H is the nonlinear operation and x_i is the ith layer feature.

The concatenation of the feature maps should be implemented correctly to ensure the DenseNet structure feasible, therefore few changes are needed to make it possible such as, down-sampling. The concatenation operation cannot be carried out if the feature maps size is continuously changing. To work out the concept of Down-sampling the dense blocks are used which contain transition layers between them. There are three Dense blocks in the DenseNet 201 structure, the input layer, transition layers and the global average pooling (GAP) layer. These transition layers have the operations like: batch normalization, convolution, and pooling. Each layer receives feature maps from all earlier layers, for achieving the easy training and increasing the efficiency of parameter, with the concatenated feature maps that are produced in the previous layer into next layers. This might also create an explosion of feature maps. In order to avoid that, the number of output maps in each layer is fixed. So, when there are a greater number of inputs, a bottleneck layer is implemented with a 1x1 convolution prior to 3x3 convolution layer to save the cost of computing and to reduce the number of feature maps. The global average pooling (GAP) which is identical to traditional pooling method is used for large feature maps reduction, as it can reduce the entire slice into single value or bit.

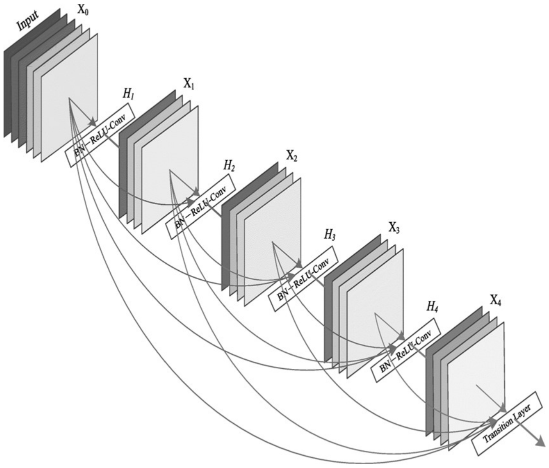


Figure:03 DenseNet Model

3.2.2 VGG19 algorithm

VGG is abbreviation of Visual Geometry Group, is also a CNN based model as in figure (04), which is upgraded from the model VGG16.Here, the existing 16 layers in VGG16 are expanded as 19 layers, among which 16 of those are convolution layers and 3 of them are fully connected. This model uses the piling of the convolutions but due to a drawback called diminishing gradient the level of piling is restricted. Hence the training of deep convolutions is quite difficult. Using this model up to 1000 object categories can be classified which is pre-trained. This model is applied in different medical diagnosis applications for classification and detection like diabetic retinopathy [14].

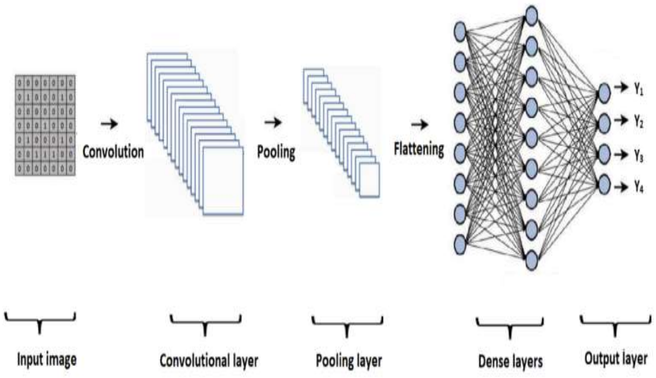


Figure 04: Basic CNN architecture

The input size for VGG19 is 224,244 consisting of 3 RGB channel as in the figure 05. This model has the convolution layers of 3x3 filter with stride 1 rather than having more hyper-parameters, also a 2x2 filter with stride 2 in max pool layer and similar padding. These layers are constantly built in the entire architecture. There are 64 filters in Conv-1, 128 filters in Conv-2, 256 filters in Conv-3 and 512 layers in Conv-4 and Conv-5 respectively, followed by three Fully Connected Layers (FCL).
The first two layers contain 4096 channels and the last one is a 1000-way classification. The final layer is the soft-max layer.

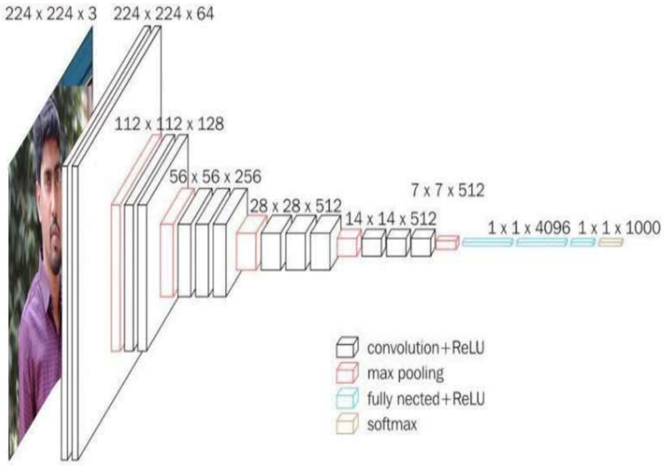


Figure 05: VGG19 layer architecture

3.2.3. Inception Resnet V2 algorithm

InceptionResNeTV2 is a CNN based model. There are 164 layers in the model, which makes it capable of classifying the images up to 1000 categories. Because of this depth of the model, this is widely used in image classification [14] along with the other model used in this work.

It is built using a mix of the Residual connection and the Inception structure. Several sized convolutional filters and residual connections are merged in the Inception-Resnet block. Using residual connections not only prevents the degradation issue brought on by deep structures but also shortens training time.

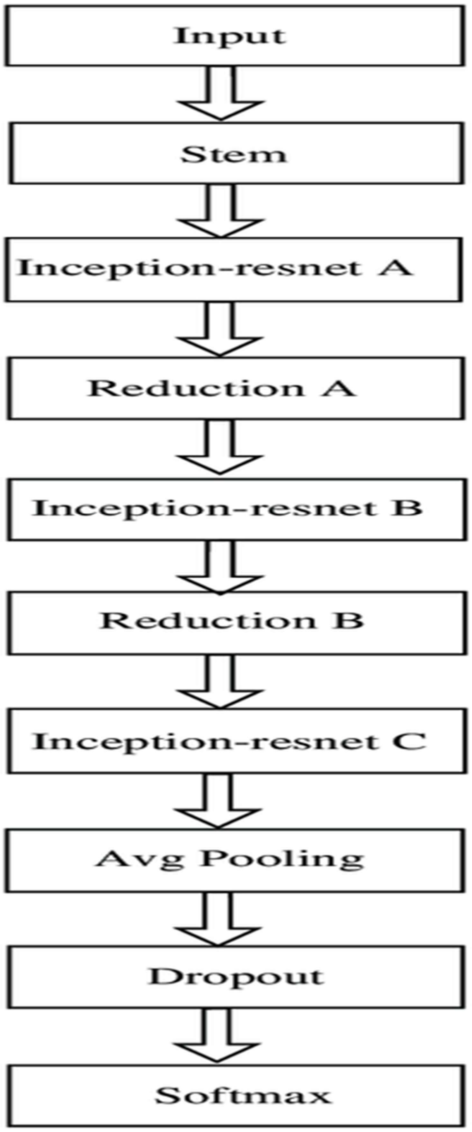


Figure 06: InceptionResNet V2

4. RESULTS

This is a summary of the experimental findings after training and testing the dataset with three distinct models: VGG19, DenseNet 201, and InceptionResNetV2. The DenseNet 201 model generated the results with the highest Accuracy (0.9986) and Validation Accuracy (0.9873). For each model, an epoch value of 20 is taken into consideration. The training accuracy for the other two models, VGG19 and InceptionResNetV2, was 0.9953 and 0.9977, respectively. The accuracy and loss using these models for training and validation are depicted in the figures [7-9]. The ratio of total accurate classifications obtained to the total number of categories evaluated is used to calculate accuracy.

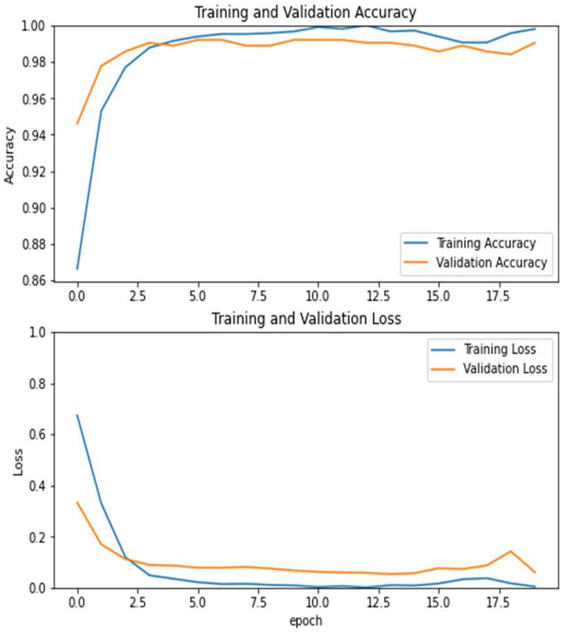


Figure 07: Training and Validation accuracy using DenseNet201

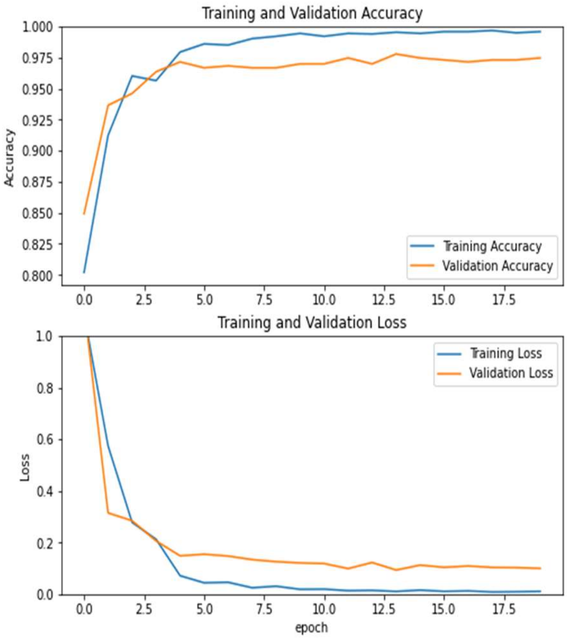


Figure 08: Training and Validation accuracy using VGG19

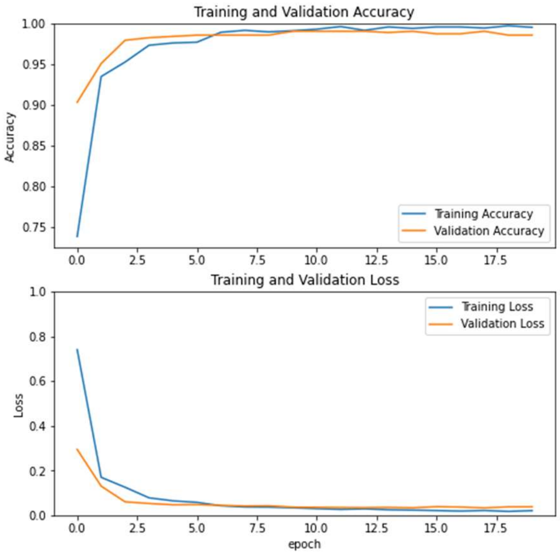
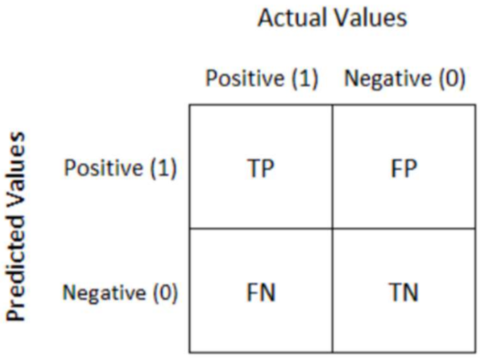


Figure 09: Training and Validation accuracy using InceptionResNetV2

For a binary classification issue with two classes or a multi-class classification problem, the confusion matrix is a summary of the projected outcomes in a particular table structure that makes it possible to see how well the machine learning model is doing (more than 2 classes)



		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 10: Confusion Matrix

Confusion matrix of a binary classification

- True Positive refers to this. It can be regarded as True because the model anticipated a positive class.
- FP stands for False Positive. While it is False, it might be viewed as the model having predicted a positive class.
- FN stands for False Negative. While it is False, it might be taken as the model's expected negative class.
- True Negative refers to this. It can be understood as True because the model expected a negative class.

Consider a diagnostic test that looks to check if a person has a certain ailment to obtain an acceptable example in a real-world issue. When a person tests positive but does not truly have the condition, it's called a false positive. On the other side, a false negative happens when a person tests negative and appears to be healthy while they have the condition.

The following are the values plotted by the three models that are used in the detection process, containing 227 images of stroke faces and 387 images of normal face. Figure 11, represents the matrix using the DenseNet 201 model, where the TP value is 227 and FP is 0. For normal faces, FN value is obtained as 1 and TN value is 386.

Figure 12, represents the matrix using the VGG19 model, where the TP value is 226 and FP is 0. For normal faces, FN value is obtained as 2 and TN value is 385

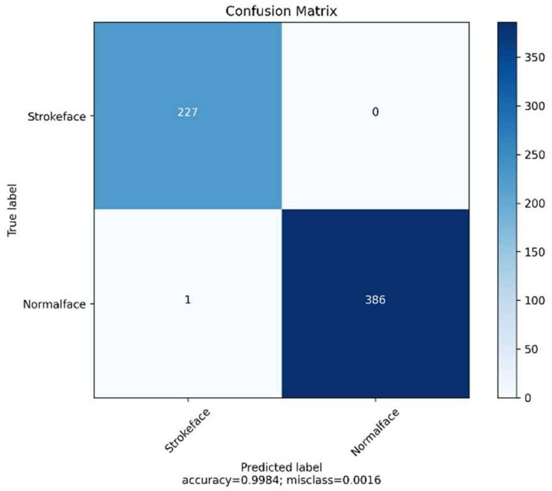


Figure 11: Confusion matrix for DenseNet 201 model

.

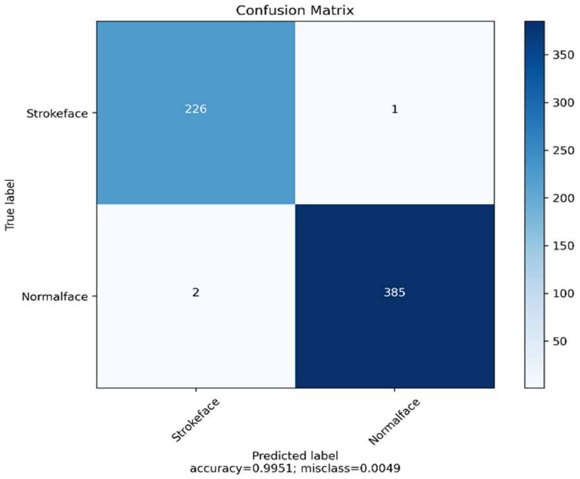


Figure 12: Confusion matrix for VGG19 model

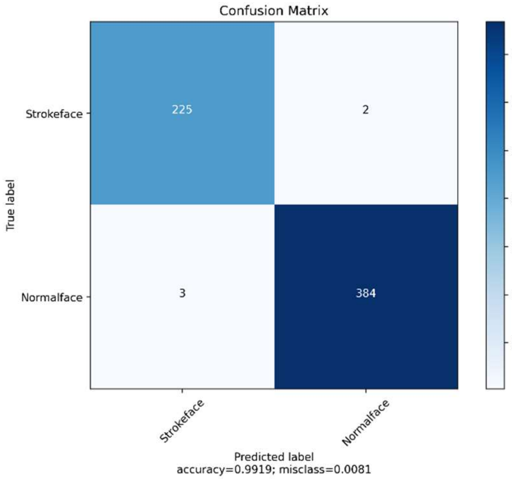


Figure 13: Confusion matrix for InceptionResNetV2 model

Figure 13, represents the matrix using the InceptionResNetV2 model, where the TP value is 225 and FP is 2. For normal faces, FN value is obtained as 3 and TN value is 384.

5. CONCLUSION

In this work, the proposed method uses the dataset containing droopy facial images are compared with the normal face images for classification. Using the best image classification models in deep learning, detection of the facial images that are infected with palsy is achieved. The trained models can detect the same with high accuracy. The comparison among the three models along with previous works, convey that DenseNet201 has the best accuracy for training as well as validation. Therefore, this approach is useful in detecting the persons with facial deformity. The above work can be used in future for image analysis for computing the level of deformation in the face by using key point analysis, which will enhance the understanding of recovery stage for patients who are having temporary facial palsy.

REFERENCES

[1] GCha CI, Hong CK, Park MS, Yeo SG. Comparison of facial nerve paralysis in adults and children. Yonsei Med J. 2008 Oct 31;49(5):725-34. doi: 10.3349/ymj.2008.49.5.725. PMID: 18972592; PMCID: PMC2615370.

[2] I. Song, N. Y. Yen, J. Vong, J. Diederich, and P. Yellowlees, ''Profiling bell's palsy based on House-Brackmann score,'' J. Artif. Intell. Soft Comput. Res., vol. 3, no. 1, pp. 41–50, Dec. 2014.

[3] Parra-Dominguez GS, Sanchez-Yanez RE, Garcia-Capulin CH. Facial Paralysis Detection on Images Using Key Point Analysis. *Applied Sciences*. 2021; 11(5):2435. https://doi.org/10.3390/app11052435

[4] Chandaliya, Rishabh & Joshi, Praveen & Afli, Haithem. (2021). TeleStroke System (TSS) - Stroke Detection using Machine Learning.

[5] G. -S. J. Hsu, W. -F. Huang and J. -H. Kang, "Hierarchical Network for Facial Palsy Detection," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 693-6936, doi: 10.1109/CVPRW.2018.00100.

[6] Barbosa J, Seo WK, Kang J. paraFaceTest: an ensemble of regression tree-based facial features extraction for efficient facial paralysis classification. BMC Med Imaging. 2019 Apr 25;19(1):30. doi: 10.1186/s12880-019-0330-8. PMID: 31023253; PMCID: PMC6485055.

[7] Jiang C, Wu J, Zhong W, Wei M, Tong J, Yu H, Wang L. Automatic Facial Paralysis Assessment via Computational Image Analysis. J Healthc Eng. 2020 Feb 8; 2020:2398542. doi: 10.1155/2020/2398542. PMID: 32089812; PMCID: PMC7031725.

[8] G. Yolcu et al., "Deep learning-based facial expression recognition for monitoring neurological disorders," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017, pp. 1652-1657, doi: 10.1109/BIBM.2017.8217907.

[9] A. M. Ridha, W. Shehieb, P. Yacoub, K. Al-Balawneh and K. Arshad, "Smart Prediction System for Facial Paralysis," 2020 7th International Conference on Electrical and Electronics Engineering (ICEEE), 2020, pp. 321-327, doi: 10.1109/ICEEE49618.2020.9102600.

[10] Z. Guo et al., "Deep assessment process: Objective assessment process for unilateral peripheral facial Paralysis via deep convolutional neural network," 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 2017, pp. 135-138, doi: 10.1109/ISBI.2017.7950486.

[11] C. Jiang et al., "Automatic Facial Paralysis Assessment via Computational Image Analysis", Journal of Healthcare Engineering, vol. 2020, pp. 1-10, 2020. Available: 10.1155/2020/2398542.

[12] ZHANG, Y.-D., PAN, C., CHEN, X. AND WANG, F. (2018). Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling. Journal of Computational Science 27, 57-68.

[13] Maneet Kaur Bohmrah, Harjot Kaur,Classification of Covid-19 patients using efficient fine-tuned deep learning DenseNet model,Global Transitions Proceedings,Volume 2, Issue 2,2021,Pages 476-483,ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2021.08.003.

[14] Sudha, V. & Ganeshbabu, Dr. (2020). A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning. Computers, Materials & Continua. 66. 827-842. 10.32604/cmc.2020.012008.

[15]Y. Bhatia, A. Bajpayee, D. Raghuvanshi and H. Mittal, "Image Captioning using Google's Inception-resnet-v2 and Recurrent Neural Network," *2019 Twelfth International Conference on Contemporary Computing (IC3)*, 2019, pp. 1-6, doi: 10.1109/IC3.2019.8844921.

[16] K. Mehta, "Facial_Droop_and_Facial_Paralysis_image," Kaggle, 23-Aug-2019. [Online]. Available: https://www.kaggle.com/kaitavmehta/facial-droop-and-facial-paralysis-image. [Accessed: 20-Jan-2022].

[17] "UTK Face Cropped", Kaggle.com, 2022. [Online]. Available: https://www.kaggle.com/abhikjha/utk-face- cropped. [Accessed: 20- Jan-2022].