Content

* Abstract
* Aknowledgement
* Introduction
* Methodology-:
  + - * Package used
      * Data
      * Data preprocessing
      * Proposed technology
* Training & Testing
* Conclusion

Abstract

Pain is naturally relative; each person interprets their feeling of pain with respect to their previous experiences with the sensation. Facial Pain expression is an important aspect for assessing pain, especially when the patient is unable to speak i.e, Baby. Facial expression is one of few alternatives available by which we can access the pain of patient when the verbal report is unavailable i.e, in the case of infants or person having verbal communication deflect. Even when verbal report is available, however, assessment of facial expression of pain brings added value. Facial action coding system (FACS) based action units (AUs) is well founded of detecting facial expression.

In this project we propose a pain recognition system that has ability to detect pain of a persons by detecting its facial expression. we have used convolutional neural network to detect, classify and give percentage of pain based on facial muscle based action unit.

This report also demonstrates the algorithm and python library we used to make this system. The project illustrates the prowess of artificial intelligence such as neural network and various python library such as matplotlib, Numpy, Keras, pandas.

**Acknowledgement**

We would like to this opportunity to thank our faculty mentor, Mrs. Divya Gupta for providing us her continuous quidance without which this project would not have been completed. We would like to extend our sincerest gratitude to the Head of Department, Dr. B.D. Mazumdar, for his constant support throughout the project. Last but clearly not the least, we would thank The Almighty for giving us strength to complete the project on time.

1. **Introduction**

International Association for the Study of Pain (IASP) defines pain as an “unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage” [1].

The detection of pain in the body is a useful indicator of problems or complications, such as injury. While the incidence of pain can be different person to person[2]. This subjectivity introduces difficulties in diagnosis and treatment of patients. The assessment of pain is a useful clinical criterion for the detection of pain some of which are unable to recognize or under recognize.

Thus the ability to accurately recognize the occurrence of pain in a clinical setting is therefore crucial to help patients. Thus, the notion of computer vision-based automatic pain level assessment was introduced.

Facial pain expression can be considered as a subset of facial expression and expression emotion valley regarding to experiencing pain[2]. It can also provide information about the severity of pain that can be assessed by using the facial action coding system[6,54]. For a long time the FACS has been used to measure facial expression appearance and intensity.

The task of assessing the pain level from facial image or video is rather challenging. A substantial body of literature has been produced in the recent years to address the challenges[3,10,30,48,50]. From the facial images in the figure, we can see that the pain and non-pain frames may not present enough visual difference. The challenges also increase in the presence of external factors like ‘smiling in pain’ phenomenon and gender difference (male’s vs female’s way of experiencing) to pain [29,55,31].This in turns result to a non-linearly wrapped facial emotion levels in a high dimensional space [53].

Recent advances in facial video analysis using deep learning frameworks such as Convolutional Neural Networks (CNN) or Deep Belief Networks (DBN) provide the notion of realizing non-linear high dimensional compositions [51]. Deep learning architectures have been widely used in face recognition [44,36,19,57], facial expression recognition [58,25], emotion detection [51,22,24]. Pain level estimation using a deep learning framework was also proposed [59]. Employing deep learning framework for pain level assessment from facial video entails two kinds of information processing from facial video sequences:

1. spatial information,
2. temporal information

Spatial information provides pain related information in the facial expressions of a single video frame. On the other hand, temporal information exhibits the relationship between pain expressions revealed in consecutive video frames.

1. **Methodology**
2. **Package Used:-**
3. **Keras:-** Keras is a python library specially written for Neural Network.It runs on top of Tensorflow or Theano. It is developed to implement deep learning model that can help AI enthusiast to use all the model in one place. It is fast as well as easy for research and development. It is written in python 2.7 as well as python 3.7 as well as can be executed on GPUs and CPUs. It is released under the permissive MIT license. Keras is developed and maintained by François Chollet, a Google engineer under the fourguiding principles known as Modularity,Minimalism,Extensibility,Python as explained below:-
4. **Modularity:** Modularity is a good programming practice and it helps to maintain reusability of code. All the components of deep learning model are discrete and can be combined and used in an arbitrarily ways.
5. **Minimalism:** This library provides all the functionality to achive an outcome and maximize readability.
6. **Extensibility:** Extensibility is a way to add new feature in the old one. Keras is developed in such a waysuch that new components are easy to add and use withing the framework intended for researchers to trial and explore new ideas.
7. **Python:** All the functionality is implemented in python.
8. **Numpy:-NumPy** is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications.
9. **Matplotlib:-Matplotlib** is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged. SciPy makes use of Matplotlib.
10. **Pandas:-**pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.
11. os, sckit-learn etc
12. **Data**

In this project we have used manual data from different sources but we can also use some well known dataset given as below:-

1. Biovid Dataset

It contains data from ninety participants from three age groups compromising thirty images of men and women, in the age categories of 18-35, 36-50 and 52-65

This dataset has five parts:

* Pain simulation (Videos)
* Pain simulation with partially occluded faces (Short windows)
* Pain simulation (Long videos)
* Posed pain and different emotions
* Emotion elicitation (Video)

1. Cohn-Kanade Dataset (CK+)

Contained data from 210 participants in the age range of 18-50 years, 69% of whom were female. It was split into emotion videos and emotion labels. The participants were asked to create twenty-three facial displays for the. All images were taken on a Panasonic AG-7500 camera.

1. **Data Preprocessing:-**

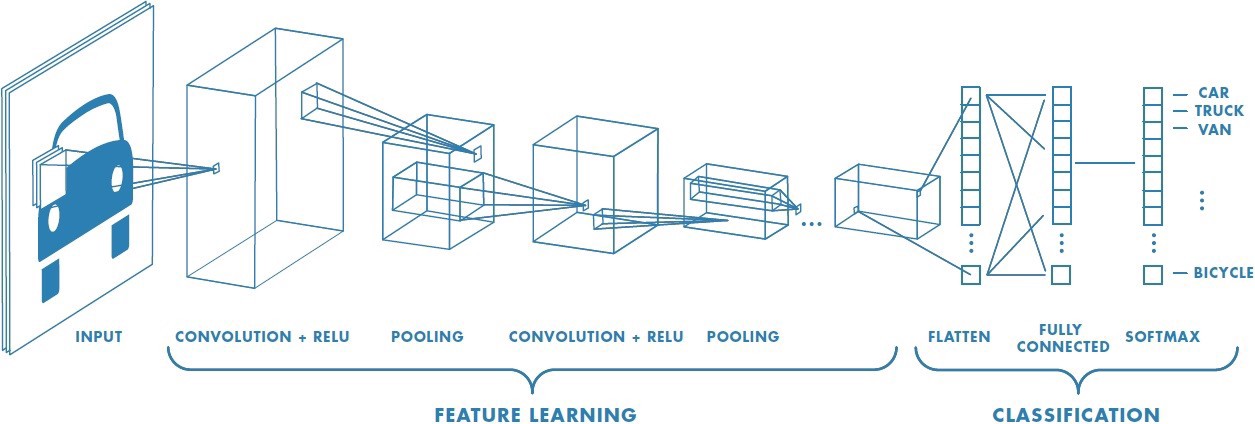
Colored images require more computation power as compared to gray images as 3 channels for RED, Green and Blue are used and thus it increases the requirement of computation power. Thus we will convert all colored images into gray images. All the images are converted into the same size so that handling the images and analysis of facial expression will be easier lots.

**V. Proposed Technology:-**

**a. Convolution Neural Network**

A Convolutional Neural Network (CNN) is a most popular Deep Learning model which takes an input image,video or sound, assign learnable weights and biases to various objects in the image,video or song and be able to classify one from the other. The convolutional Neural Network requires much less pre-processing and other steps as compared to other classification

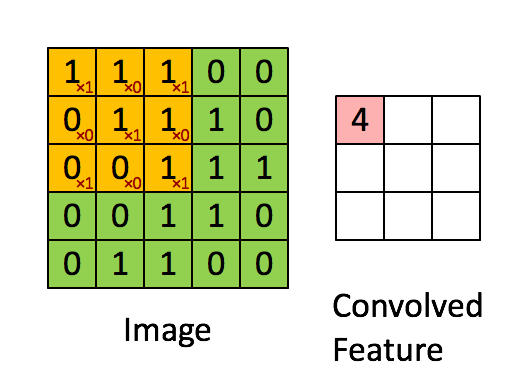
algorithms. In primitive algorithms, filters are manually-engineered which requires lot of training while CNN have the ability to learn these filters automatically. The CNN architecture is similar to that of the connectivity pattern of biological Neurons in the human brain and it is the network that makes Neurons a much powerful. A neural Network recieves signals as an electric pulse from other neurons throught synapses. A human brain which is analyzing visuals are is much more powerfull than 1000 supercomputers and much less powerful as a one-tenth of a pocket calculator in number calculations.



Rather than using labels to learn and classify an image, convolutional neuralnetwork (CNN) usesanimages pixels to derive a difference/similarity between images. A CNN has layers like a normal neural network but has a convolutional layer in between those that use a set of learnable filters to detect features etc and produces an activation map which is passed to the next layer in the neural network (Udofia, 2018). Finally, a neural network uses an activation function to decide whether to ‘fire’ a neutron, i.e. make a decision.Finally, a neural network uses an activation function to decide whether to ‘fire’ a neutron, i.e. make a decision. Spatial and temporal dependencies in an image,video or sounds are very important for classification and those cap be successfully captured throught ConvNet through the application of relevant filters.

As number of parameters are reduced and usability of weights increases, The convNet architecture performs a better fitting to the video,sound or image dataset. In other words, the ConvNet can be trained to understand the complex features of the image better.

**Convolution Layer — Kernel**

e.g. Image Dimension = 5 (H) x 5 (B) x 1 (Number of channels, eg. RGB) H=Height,B=Breadth

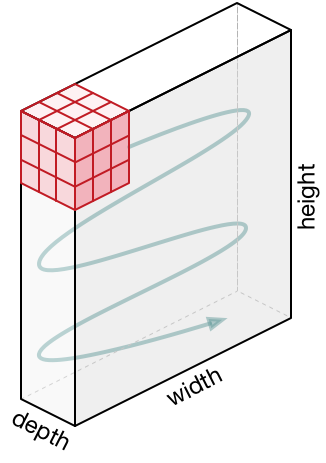
In the above, the squared section resembles **5x5x1 input image, I**. The squared window which moves into the image matrix and extract the required features in the first step of a convolutional layer is called the **Kernel/Filter, K**, In the small squared region represented in the color yellow. We have selected **a kernel as a 3x3x1 matrix.**

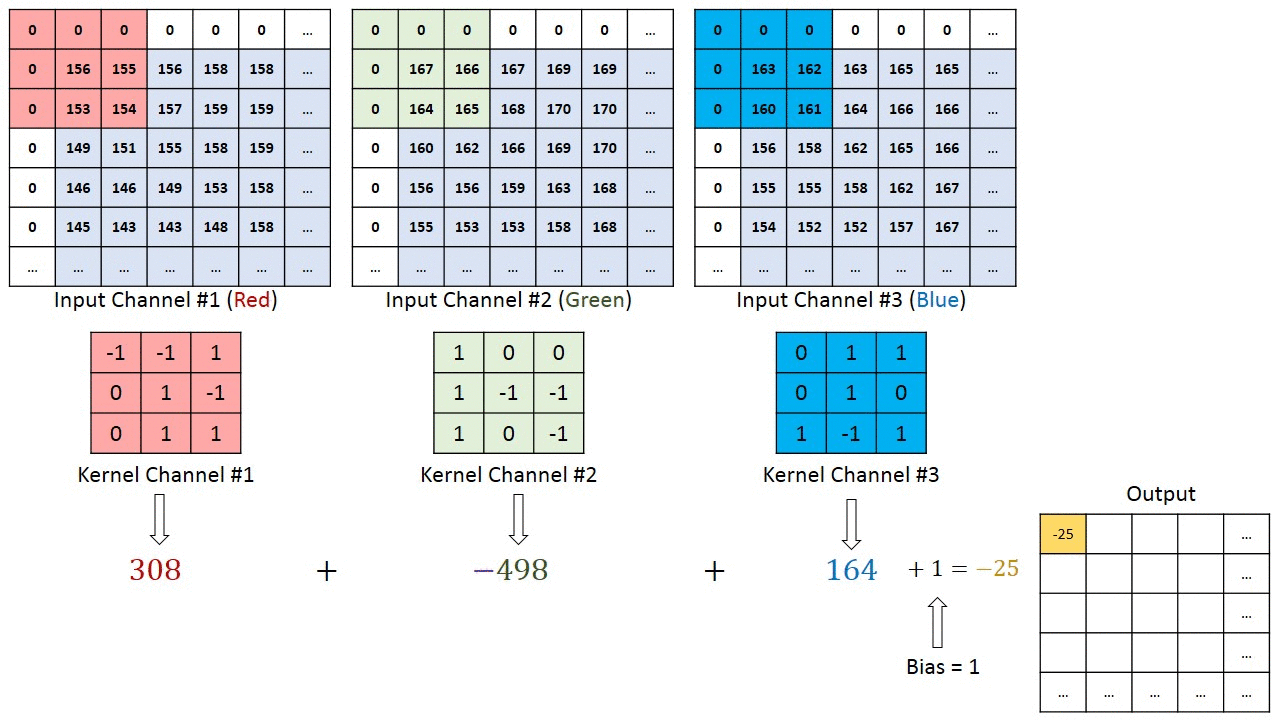
Kernel/Filter, K = 1 0 1

0 1 0

1 0 1

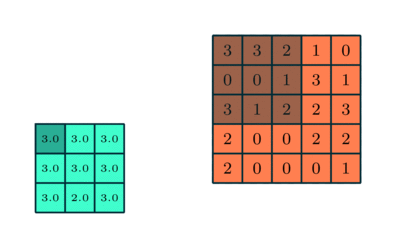
The 3x3 matrix Kernel moves 9 times because **Stride Length is 1** which is **Non-Strided** and each time performing a **matrix mult operation between Kernel and the region P of the image(5x5)**.



The kernel visits to the right at first with a fixed stride value till the complete width and moving to the down from left of the image with the same stride value as chosen before. It repeats the same process until the whole image is visited.

If we are using images having multiple channels (e.g rgb), the kernal depth is same as depth of input image. Matrix muliplication is operated between kn and in stack([K1, I1]; [K2, I2]; [K3, I3]) and all the results are added with the bias to give us a squashed one-depth channer convoluted feature output.

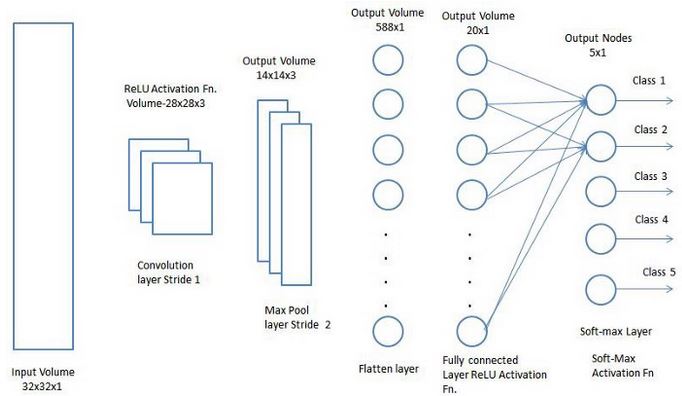
Pooling layer



As in the convolutional layer, To reduce the spatial size of the convoluted feature, pooling layer is used. It is used to decrease the computational power required to process the large data with the dimensionality reduction technique. **It extracts strong or dominant features that are positional and invariant** and thus it trains the model effectively in less effort.An ith-convolutional layer is formed withthe convolutional layer and pooling layer.if the image is more complex, colvolutional layer can be increased to capture the low level details of the image but it will requires more computational power.Thus with the about process, we have made the model to understand the crucial features successfully. After this process, we will give this to a regular neural network for the classfication purpose.

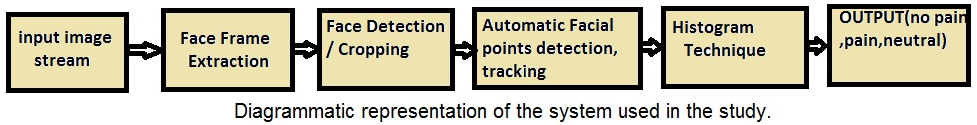


## Classification — Fully Connected Layer (FC Layer)



we can embed one fully connected layer that is a cheap way to learn the non linear combinations of features at the convolutional layer.The fully-connected layer is enough to learn all the complex features that are most important for classfication task.As we have converted out input image into a well suitable form for the CNN ,We shall flatten the input image into the column vector.At every iteration of training,flatten output is feeded to the feed-forward neural network and back propogation is applied to adjust the learnable weights to every link between neurons.Over a series of epochs of training, the CNN model is well vershed to distinct the important and domination features to images and by using the softmax classfier we classify into the corresponding.As per the requirements ,Many architectuers of Convolutional Neural network have been developed which are keys to develop many Deep learning algorithm. Some popular CNN architecture are listed below:-LeNet

1. AlexNet
2. VGGNet
3. GoogLeNet
4. ResNet
5. ZFNet
6. **System Diagram**



**Step 1.**Input the Stored image to the System. These images may be RGB image of png or jpeg format.

**Step 2.** Convert the RGB image to gray imageStep 3. Face detection and cropping is performed and it resutls a 64\*64 matrix.

**Step 3.** (a).Track the facial landmark in an image in a sequence.

(b). The facial landmarks are tracked in the previous image, in current image it is referred as initial landmarks.

(c). A hybrid model was created to detect and track the initial landmarks.

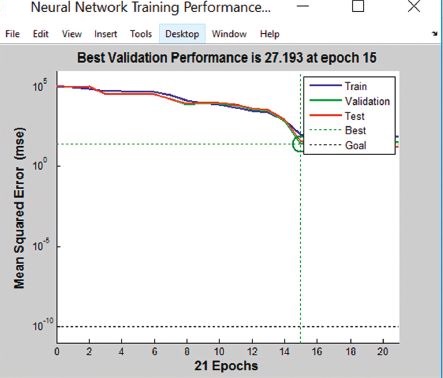
(d). It outputs well-aligned landmark coordinated, the information of head pose, and the visibility of each landmark.

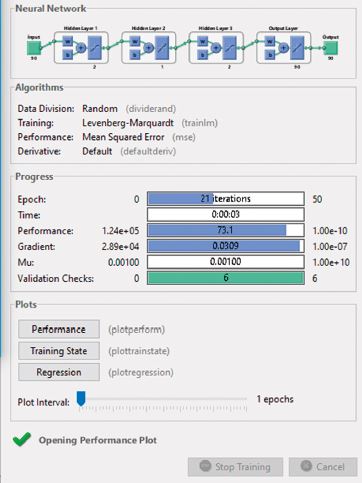
**Step 4.** Extraction of feature points or Facial Landmark Points.In this approach, there are 6 different features that are extracted names as- Scale,Pitch,yaw,roll,shift of x coordinates from referenced shape,shift of y coordinates from the reference shape along with local binary patterns.

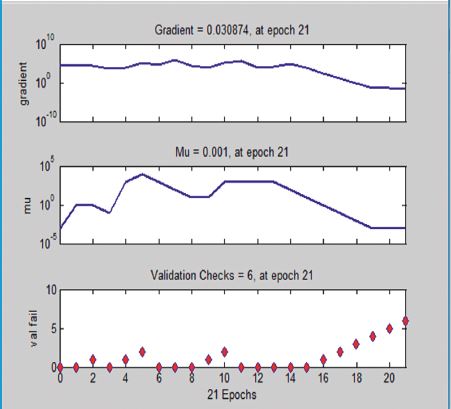
**Step 5.** Train the model with reference to database.

**Step 6.** For testing , apply colour conversion, facedetection, and cropping and localise facial landmarks in a facialframe.

Training and Testing with Multilayer Back propagation Neural Network (MLBPNN).

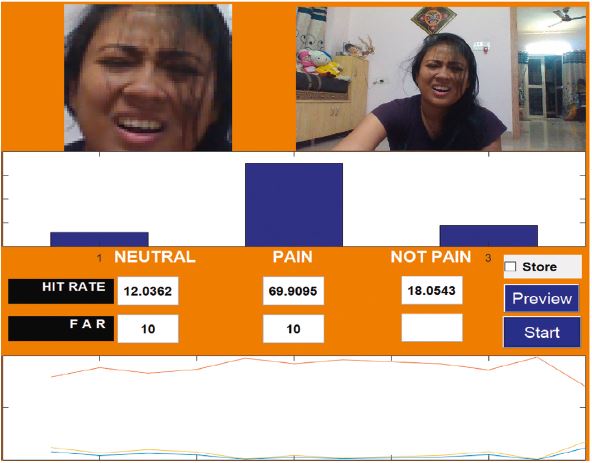
Despite the given training time, For adaptive learning the neural network technique is always a good fit. It is also best suitable for facial expression and emotion recognition.MLBPNN is best way to detect pain and to compare the accuracy and the processing speed of the proposed algorithm.MLBPNN contains one input layer, one output layer and three middle hidden layers with five nodes.Two nodes are used in the first hidden layer and one node in second and again two nodes in third hidden layer.During the training period, image by image results is taken divide them into three results ( pain face, neutral face, and no pain face). We are using three hidden layers that increases training time and more computational power and thus capture the low-level feature.In below images we have given an overview of training performance, validation performance and training state.











Conclusion

Using content based filtering for a Facial PainExpression, despite its flaws, remain a viable focuses more on employed at initial stages of a facial expression system.

Using facial landmarks to identify emotions is not an accurate way to identify an emotion, because it only took in to consideration the movement of the face, and more is needed to identify specific emotion, including body movement, heart rate and other biological signals. As such a system to monitor and identify an emotion cannot be based purely on facial signals.

Aggregating this model and the other biological signals such as voice we can more accurately describe the pain of the patient.