COMPUTERIZED FACIAL FEATURES RECOGNITION USING FUNCTION EXTRACTION

Dr V Srilakshmi[1], Associate Professor, Department of Computer Science and Engineering, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India
Y Siva Prasad[2], Assistant Professor, Department of Computer Science and Engineering, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India
R Leela[3], UG Student, Department of Computer Science and Engineering, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India
D Venkata Ganesh[4], UG Student, Department of Computer Science and Engineering, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India
G Hari Priya[5], UG Student, Department of Computer Science and Engineering, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India
S Yashwanth Chowdary[6], UG Student, Department of Computer Science and Engineering, DVR & Dr. HS MIC College of Technology, Kanchikacherla, Andhra Pradesh, India

Abstract

Human emotions are spontaneous psycho-emotional states. The lack of a clear link between emotions and facial expressions and the considerable variability makes facial recognition a difficult area of research. Pattern recognition considers features such as histograms of oriented gradients (HOG) and scale-invariant feature transforms (SIFT). These features are extracted from the image according to a manually defined algorithm. In recent years, machine learning (ML) and neural networks (NN) have been used for emotion recognition. This report uses Convolutional Neural Networks (CNN) to extract features from images and detect emotions. We used the Python dlib toolkit to detect and extract 64 key facial landmarks. Training a CNN model using grayscale images from the FER-2013 dataset, he classified utterances into five emotions: happy, sad, neutral, fear, and anger. This article aims to identify basic human emotions. This paper aims to identify basic human emotions . The facial emotions such as happy, sad, angry, fear, surprised, neutral emotions are considered as basic emotions.

1.INTRODUCTION

Facial emotions are important factors in human communication that help to understand the intentions of others. In general, people infer the emotional state of other people, such as joy, sadness and anger, using facial expressions and vocal tones. Facial expressions are one of the main information channels in interpersonal communication. Therefore, it is natural that facial emotion research has gained a lot of attention over the past decade with applications in perceptual and cognitive sciences. Interest in automatic Facial Emotion Recognition (FER) has also been increasing recently with the rapid development of Artificial Intelligent (AI) techniques. They are now used in many applications and their exposure to humans is increasing. To improve Human Computer Interaction (HCI) and make it more natural, machines must be provided with the capability to understand the surrounding environment, especially the intentions of humans. Machines can capture their environment state through cameras and sensors. In recent years, Deep Learning (DL) algorithms have proven to be very successful in capturing environment states. Emotion detection is necessary for machines to better serve their purpose since they deliver information about the inner state of humans. A machine can usea sequence of facial images with DL techniques to determine human emotions.

2. MOTIVATION

AI and machine learning (ML) are frequently used in a variety of fields. They have been applied to data mining to find insurance fraud. Data mining techniques based on clustering were employed in to find trends in stock market data. FER, Electroencephalography (EEG), and spam detection are a few examples of pattern recognition and classification situations where ML algorithms have been particularly useful. Cost-effective, dependable, and quick FER solutions can be provided using ML.

DOI: 10.36893.JK.2023.V13I04N16.0090-0096

3.FACIAL EMOTION RECOGNITION

FER usually consists of four phases. Drawing a rectangle around a face in an image after identifying it is the first stage. The next is to look for landmarks within the face region. The third stage is to separate the facial components' spatial and temporal properties. The recognition results are produced in the final step by using a Feature Extraction (FE) classifier and the extracted features. The FER process for an input image with a face region and facial landmarks detected is shown in Figure 1. Facial landmarks are visually noticeable places on the face, such as the tip of the nose, the corners of the mouth, and the ends of the brows, as seen in Figure 2. Features include the local texture of a landmark or the pair-wise placements of two landmark points. The descriptions of 64 primary and minor landmarks are provided in Table 1. Using pattern classifiers, the face's spatial and temporal features are retrieved, and the expression is calculated based on one of the facial categories.

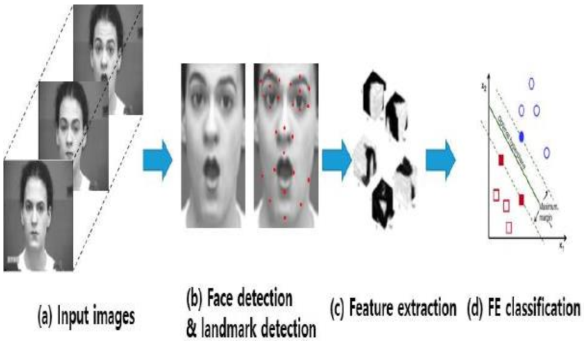


Figure1 FERproceduresforanimage

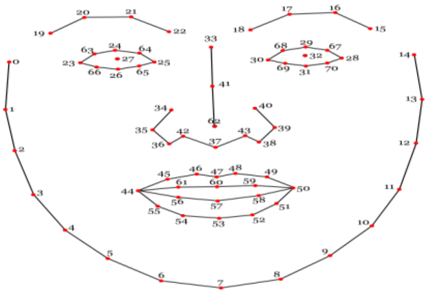
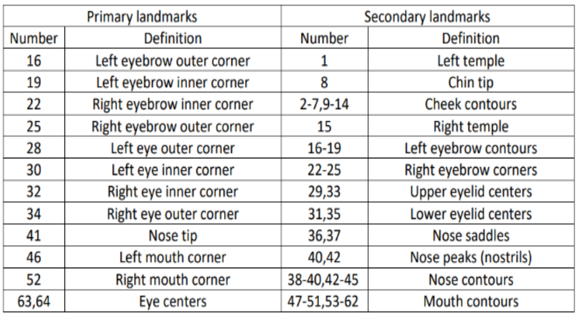


Figure 2Faciallandmarkstobeextractedfromaface.



Primary landmarks		Secondary landmarks	
Number	Definition	Number	Definition
16	Left eyebrow outer corner	1	Left temple
19	Left eyebrow inner corner	8	Chin tip
22	Right eyebrow inner corner	2-7,9-14	Cheek contours
25	Right eyebrow outer corner	15	Right temple
28	Left eye outer corner	16-19	Left eyebrow contours
30	Left eye inner corner	22-25	Right eyebrow corners
32	Right eye inner corner	29,33	Upper eyelid centers
34	Right eye outer corner	31,35	Lower eyelid centers
41	Nose tip	36,37	Nose saddles
46	Left mouth corner	40,42	Nose peaks (nostrils)
52	Right mouth corner	38-40,42-45	Nose contours
63,64	Eye centers	47-51,53-62	Mouth contours

Table1Definitionsof64primaryandsecondarylandmarks

By enabling end-to-end learning directly from the input images, DL based FER systems significantly minimise the dependence on face-physics based models and other preprocessing techniques. Convolutional Neural Networks (CNNs) are the most often used DL model. With a CNN, a feature map is created by filtering an input image via convolutional layers. The output of the FE classifier is then passed to fully connected layers, which identify the facial expression as belonging to a class.

The dataset used for this model is the Facial Emotion Recognition2013 (FER2013) dataset. This is an open source dataset that was created for a project then sharedpublicly for a Kagglecompetition. It consists of 35,000 grayscale size 48 × 48 face images with various emotionlabels.Forthisproject,five emotionsareused,namelyhappy,angry,neutral,sadandfear.

4.PROPOSED METHOD

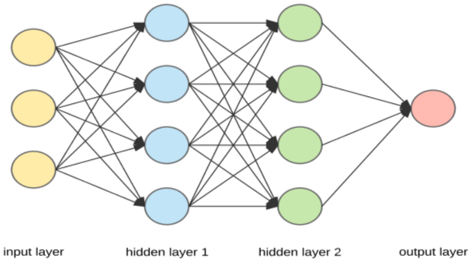
This paper aims to identify basic human emotions like happiness, sadness, anger, fear, surprise, and neutral emotions. Here is a real-time facial emotion recognition system based on the CNN model. It provides significant, accurate object detection and extracts high-level features that help achieve tremendous performance in classifying the image and detecting objects. This model provides a more accurate result than other methods because of the large number of hidden layers and cross-validation in the neural network.

5.CONVOLUTIONALNEURAL NETWORKS (CNN)

Convolutional neural networks, often known as convents or CNNs, are a type of machine learning. It is one of the many different kinds of artificial neural networks that are applied to numerous applications and data types. Specifically used for image recognition and other tasks involving the processing of pixel data, CNNs are a type of network design for deep learning algorithms. Although there are different kinds of deep learning neural networks, CNNs are the preferred network architecture for detecting and classifying objects. They are therefore ideal for computer vision (CV) activities and for scenarios where accurate object recognition is crucial, such as in autonomous vehicles and facial recognition.

6.INSIDE CONVOLUTIONAL NEURAL NETWORKS

An essential component of deep learning techniques are artificial neural networks (ANNs). Recurrent neural networks (RNNs), one type of ANN, take input from time series or sequential data. It is appropriate for applications involving speech recognition, language translation, natural language processing (NLP), and image captioning.

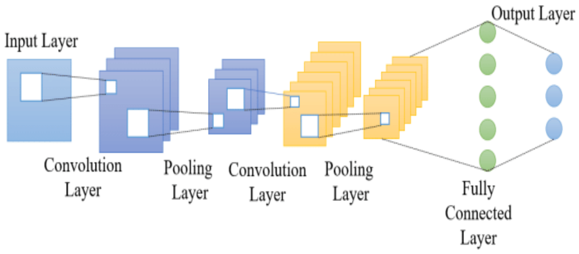


A different kind of neural network called a CNN may find important information in both time series and picture data. This makes it very beneficial for applications involving images, such as pattern recognition, object classification, and picture identification. A CNN makes use of linear algebraic concepts, including matrix multiplication, to find patterns in an image. CNNs may categories audio and signal data as well.

The structure of a CNN is comparable to the connection structure of the human brain. Similar to how the brain has billions of neurons, CNNs also have neurons, but they are structured differently. In actuality, a CNN's neurons are set up similarly to the frontal lobe of the brain, which processes visual stimuli. This configuration guarantees that the full visual field is covered, avoiding the issue with typical neural networks' piecemeal image processing that requires images to be given to them in low-resolution chunks. A CNN performs better with image inputs and voice or audio signal inputs compared to the earlier networks.

7.CNN LAYERS

A convolutional layer, a pooling layer, and a fully connected (FC) layer make up a deep learning CNN. The first layer is the convolutional layer, while the final layer is the FC layer.The complexity of the CNN grows from the convolutional layer to the FC layer. The CNN is able to identify increasingly larger and more intricate aspects of an image until it successfully recognizes the complete thing as a result of therising complexity



.

Convolutional layer: Convolutional layer, the fundamental constituent of a CNN, is where the majority of computations take place. After the first convolutional layer, a second convolutional layer may come. In order to perform convolution, a kernel or filter inside this layer must move across the image's receptive fields and determine whether a feature is present. The kernel iteratively moves throughout the entire image. A dot product is calculated between the input pixels and the filter after each cycle. A feature map or convolved feature is the result of the dots being added together. In this layer, the image is ultimately transformed into numerical values, enabling the CNN to understand the image and draw out pertinent patterns from it.

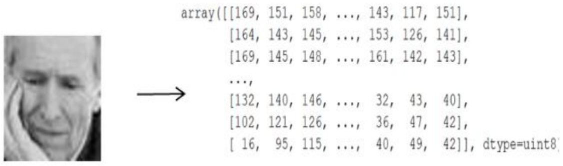
Pooling layer: The pooling layer also applies a kernel or filter to the input image, just like the convolutional layer does. Nevertheless, in contrast to the convolutional layer, the pooling layer reduces the number of input parameters and also causes some information loss. On the plus side, this layer streamlines operations and boosts CNN effectiveness.

Fully connected layer:Based on the features extracted in the preceding layers, picture categorization in the CNN takes place in the FC layer. Fully connected in this context means that every activation unit or node of the subsequent layer is connected to every input or node from the preceding layer.

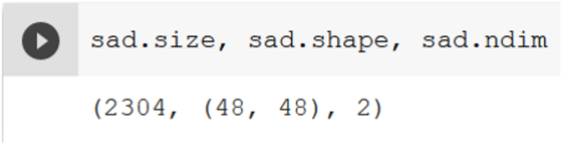
The CNN does not have all of its layers fully connected because that would create an excessively dense network. It would cost a lot to compute, increase losses, and have an impact on output quality.

8.IMAGE TO ARRAY

arrays from images Values (numbers) that represent the pixel intensities make up an image. To acquire an image from an array and transform it into a picture, utilize the array module in NumPy (nd.array). attributes. Figure displays a sad class image from the FER 2013 dataset as a NumPy array. The attributes of this image, which are 2304 pixels, 2 dimensions, and a size of 48 48 pixels, are displayed in Figure.



```
array([[169, 151, 158, ..., 143, 117, 151],
       [164, 143, 145, ..., 153, 126, 141],
       [169, 145, 148, ..., 161, 142, 143],
       ...,
       [132, 140, 146, ...,  32,  43,  40],
       [102, 121, 126, ...,  36,  47,  42],
       [ 16,  95, 115, ...,  40,  49,  42]], dtype=uint8
```



```
sad.size, sad.shape, sad.ndim

(2304, (48, 48), 2)
```

9.CONVOLUTION OPERATION

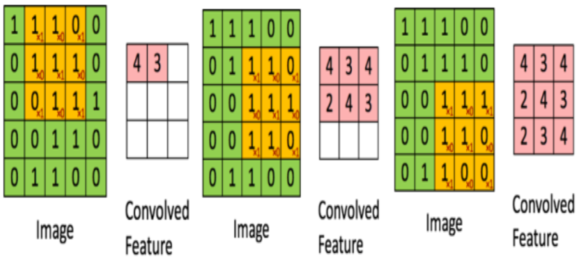
Convolution is a process that takes an input image and extracts high level characteristics like edges. The convolution layer performs the following tasks.

- The first convolutional layer (or layers) picks up information including edges, color, gradient orientation, and basic textures.
- More complex textures and patterns are learned by the subsequent convolutional layer(s).
- The final convolutional layer (or layers) picks up on features like objects or pieces of objects.

The kernel is the component that performs the convolution action. A kernel isolates the pertinent data and filters out everything else that is irrelevant to the feature map. Until it has parsed the entire width, the filter travels to the right with a specific stride length. Once the entire picture has been traversed, it returns to the left of the image with the same stride length and continues the procedure.

DOI: 10.36893.JK.2023.V13I04N16.0090-0096

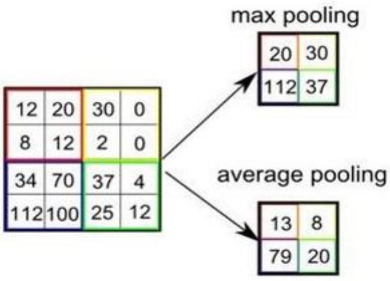
Figure displays a 5 x 5 image (shown in green) and the subsequent 3 x 3 kernel filter. The stride length is chosen as one so the kernel shifts nine times, each time performing a matrix multiplication of thekernelandtheportion ofthe image underit.



The input or kernel dimensions may be the same as those of the convolved feature. The same or appropriate padding is used for this. When the convolved feature has the dimensions of the input picture, it is said to have the same padding, and when it has the dimensions of the kernel, it has valid padding.

10.POOLING OPERATION

The spatial size of a convolved feature is decreased by the pooling layer. As a result, fewer computations are needed to analyses the data and extract the most important, rotation- and position-invariant properties. There are two different types of pooling: average pooling and maximum pooling. The largest value from the area of the picture covered by the kernel is returned by max pooling, whereas the average of the corresponding values is returned by average pooling. Figure displays the results of applying max and average pooling on an image.



11.COMPILING THE MODEL

Optimizer and metrics are the two parameters needed to compile the model. Adam is the optimizer in use. A DL model's weights are updated using the optimizer based on the loss. Accuracy, categorical cross-entropy loss, precision, recall, and F-score are the measures employed. These measures are described further below.

12.TRAINING THE MODEL

The "fer2013.csv" file is subjected to the train-test function in order to train the model. The dataset is divided into training and testing sets using this function. Testing does not make use of the training data. When the training to validation ratio is 0.80, training will use 80% of the dataset and validation will use the remaining 20%. The model weights' calculation speed is controlled by the training process's customizable Learning Rate (LR) parameter. A lower LR may result in more accurate weights (up to convergence), but it requires more calculation time. A high LR may cause the model to converge too rapidly. A dataset is run through the NN both forward and backward in the number of epochs.

13.RESULTS AND DISCUSSION

The measures that are used to assess model performance are defined in this chapter. Finally, using the training data, each model's ideal parameter values are found. These numbers are used to assess the CNN model's accuracy and loss. The model's outcomes are then discussed.

14.EVALUATION METRICS

Accuracy,loss,precision,recallandF-scorearethemetricsusedtomeasuremodelperformance.Thesemetrics aredefined below.

Accuracy:Accuracyisgivenby

$$\text{Accuracy}=\frac{\text{Numberofcorrectpredictions}}{\text{Totalnumberofpredictions}}$$

Loss:Categoricalcross-entropyisusedasthelossfunctionandisgivenby

$$\text{Loss} = -\sum_{c=1}^{m}(y_{o,c}\log(p_{o,c}))$$

where y is a binary indicator (0 or 1), p is the predicted probability and m is the number ofclasses(happy,sad,neutral,fear,angry)

15.CONFUSION MATRIX

The confusion matrix provides values for the four combinations of true andpredicted values, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative(FN).Precision,recallandF-scorearecalculatedusingTP,FP,TN,FN.TPisthecorrectpredictionofanemotion,FPistheincorrectpredictionofanemotion,TNisthecorrectprediction of an incorrect emotion and FN is the incorrect prediction of an incorrect emotion.Consider an image from the happy class. The confusion matrix for this example is shown inFigure 4.1. The red section has the TP value as the happy image is predicted to be happy. Thebluesection hasFP values asthe image ispredictedto besad,angry, neutral or fear.The yellow section has TN values as the image is not sad, angry, neutral or fear but the model predicted this. The green section has FN values as the image is not happy but was predicted to be happy.

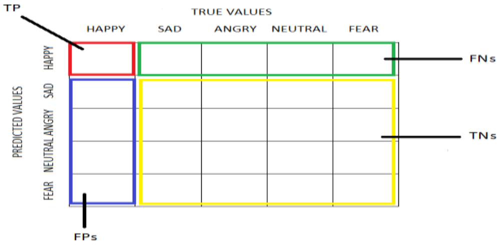


Figure 4.1 Confusion matrix for five emotions.

Recall: Recall is given by

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

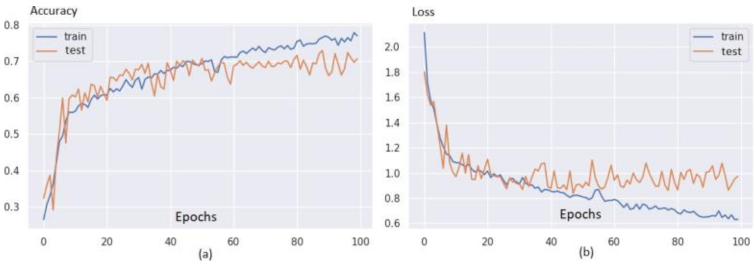
Precision: Precision is given by

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

F-score:F-scoreistheharmonicmeanofrecallandprecisionandisgivenby

$$\text{F-score} = \frac{2\times\text{Recall}\times\text{Precision}}{\text{Recall}+\text{Precision}}$$

16.RESULT



			Precision					Recall					F-score				
Trial	Acc	Loss	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
1	0.73	0.93	0.80	0.79	0.60	0.90	0.58	0.48	0.76	0.67	0.90	0.67	0.60	0.77	0.73	0.90	0.62
2	0.70	0.87	0.76	0.68	0.63	0.83	0.55	0.41	0.80	0.51	0.90	0.67	0.53	0.73	0.56	0.87	0.60
3	0.72	0.83	0.75	0.81	0.58	0.87	0.57	0.54	0.71	0.67	0.86	0.64	0.63	0.75	0.62	0.87	0.60
4	0.71	0.85	0.66	0.75	0.60	0.90	0.56	0.59	0.71	0.65	0.83	0.66	0.62	0.73	0.62	0.86	0.60
5	0.72	0.82	0.70	0.78	0.57	0.89	0.60	0.58	0.76	0.74	0.86	0.54	0.63	0.77	0.64	0.87	0.57
6	0.72	0.86	0.69	0.72	0.63	0.87	0.56	0.49	0.76	0.60	0.91	0.63	0.58	0.74	0.61	0.89	0.59
7	0.72	0.88	0.76	0.81	0.60	0.86	0.56	0.49	0.75	0.54	0.91	0.72	0.60	0.78	0.57	0.88	0.63
8	0.72	0.85	0.75	0.73	0.62	0.87	0.57	0.51	0.76	0.55	0.91	0.69	0.60	0.74	0.58	0.89	0.63
9	0.72	0.80	0.66	0.76	0.58	0.91	0.59	0.55	0.65	0.72	0.86	0.62	0.60	0.70	0.64	0.89	0.61
10	0.71	0.83	0.75	0.74	0.66	0.86	0.53	0.51	0.72	0.60	0.91	0.64	0.60	0.73	0.63	0.88	0.58
Avg	**0.72**	**0.85**	**0.73**	**0.76**	**0.60**	**0.88**	**0.57**	**0.51**	**0.74**	**0.63**	**0.88**	**0.65**	**0.60**	**0.74**	**0.62**	**0.88**	**0.60**

DOI: 10.36893.JK.2023.V13I04N16.0090-0096

In the above table, some data results are shown. Based on the total result of the data train-test, the accuracy of the model is **72%**.

CONCLUSION

In this study, a CNN model was created to identify emotions and extract facial features. 28709columns from the FER 2013 dataset were chosen to represent each of the five emotions. The emotions taken into consideration were fear, neutral, furious, sad, and pleased. Landmark features were recognized and retrieved once the camera is started from face into NumPy arrays. Convolution, pooling, batch normalization, and dropout layers were used in the first three phases of a CNN model that was created in four stages. Layers that are flattened, dense, and output make up the final step. Comparing with the exited model this proposed model gives accurate result with 72% accuracy.

REFERENCES

K. Kaulard, D.W. Cunningham, H.H. Bulthoff, C. Wallraven, The MPI facial expression database: Avalidated database of emotional and conversational facial expressions, PLoS One, vol. 7, no. 3,art.e32321, (2012).

M. Xie, Development of artificial intelligence and effects on financial system, Journal of Physics:Conference Series1187, art. 032084, (2019)

A.Nandi, F. Xhafa, L. Subirats, S. Fort, Real time emotion classification using electroencephalog-ram datastreamin e-learningcontexts,Sensors,vol.21, no.5,art.1589,(2021)

A.Raheel, M. Majid, S.M. Anwar, M. Alnowami, Physiological sensors based emotion recognitionwhile experiencingtactileenhancedmultimedia,Sensors,vol.20, no.14,art. 04037,(2020)

B.C. Ko, A Brief review of facial emotion recognition based on visual information, Sensors, vol.18, no.2,art.401,(2018)

R. Walecki, O. Rudovic, V. Pavlovic, B. Schuller, M. Pantic, Deep structured learning for facialaction unit intensity estimation, IEEE Conference on Computer Vision and Pattern Recognition,pp.5709-5718,(2017).

S.E. Kahou, V. Michalsk, K. Konda, Recurrent neural networks for emotion recognition in video,InternationalConferenceonMultimodal Interaction,pp. 467-474,(2015).

D.H.Kim,W.Baddar,J.Jang,Y.M.Ro,Multiobjectivebasedspatiotemporalfeaturerepresentationlearning robusttoexpressionintensityvariationsforfacialexpressionrecognition,IEEETransactionsonAffective Computing,vol.10,no.2,pp.223-236, (2019).

M. Liu, S. Li, S. Shan, X. Chen, AU-inspired deep networks for facial expression feature learning,Neurocomputing,vol. 159,no.1,pp. 126-136, (2015).

A. Mollahosseini, D. Chan, M.H. Mahoor, Going deeper in facial expression recognition usingdeepneural networks,IEEEWinterConferenceonApplicationsof ComputerVision,(2016).

A. Sinha, R.P. Aneesh, Real time facial emotion recognition using deep learning, InternationalJournalof Innovations&ImplementationsinEngineering, vol.1, pp.1-5,(2019).

D. Clevert, T. Unterthiner, S.Hochreiter, Fast and accurate deep network learning by exponentiallinearunits, InternationalConferenceonLearningRepresentations,(2016).

K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human level performanceon imagenet classification, IEEE International Conference on Computer Vision, pp. 1026-1034,2015.

A.Amini, A.Soleimany, MITdeeplearningopenaccesscourse6.S191,availableonline:http://introtodeeplearning.com/ (2020).

S.Shah,Acomprehensiveguidetoconvolutionalneuralnetworks,availableonline:https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53,(2018).

DOI: 10.36893.JK.2023.V13I04N16.0090-0096