International Journal of Advanced Trends in Computer Science and Engineering

Face Emotion Identification System for Visually Challenged Persons using Machine Learning

D. Gomathi[1], Dr. M. Ramalingam[2], Dr.G.Nalinipriya[3], N.Kanagavalli[4], D. Saranya[5]

[1]Assistant Professor, Information Technology, Mailam Engineering College, Mailam. dgomathi37@gmail.com.

[2] Professor, Information Technology, Mailam Engineering College, Mailam. rammalingamm_2005@yahoo.co.in

[3]Professor, Information Technology, Saveetha Engineering College, Chennai. nalini.anbu@gmail.com

[4]Assistant Professor, Computer Science & Engineering, Rajalakshmi Institute of Technology, Chennai.
kvalli.818@gmail.com

[5]Assistant Professor, Computer Science and Engineering, Mailam Engineering College, Mailam.
dsaranya93@gmail.com

ABSTARCT

In recent years, Machine Learning has regularly grown to be fast research, however, humans have an exclusive method of expressing their emotions and beneath the impact of clarity, heritage, and different factors, and there are a few problems in facial features recognition. In this paper, mainly focusing the visually challenged persons is struggling to identify the emotions of the people in front of us, due to the lack of non-visual information in their nearby environment. The proposed system is to collect the face features to checking the population's influence on the results of emotion reputation like body movements or voice indicators and additionally video series further to photograph in the emotion reputation system. The face features are extracted by using the Haar classifier which detects the face and then obtains the characteristic value. Six kinds of specific emotional categories are obtained through the Neural Network classifier training. The proposed experiment result is taken from the analysis and achieves reasonable accuracy, dependent on the test set, and test emotions.

Key words : CNN, Haar Classification, KLT algorithm, deep learning, face emotion recognition.

1. INTRODUCTION

Over the previous year, human-robot communication has gained growing attention from researchers in various fields of expertise. In this paper, we discuss one of the computer vision tasks, which are involved in creating such an integrated framework. In specific, we are concerned with identifying the presence and monitoring of individual users' attention. An assistant robot, which will have to execute multiple actions, contains future applications. When humans enter or leave the computer, interactive sports, commercials, etc. [1]. In the last

few year's identifications of artifacts has been a great challenge. Recognition of the individual face is an important research subject in computer visualization. For many computer applications, such as HCI, surveillance, human-robot interaction, etc., it is important. Facial detection has seen further development in the use of protection and safety technologies in this area to identify different circumstances. It is possible to use this monitoring domain to monitor or interact with robots.

The identification of facial features in a video is a big obstacle. The point of view, ambient brightness, and separate illuminations can be identical to these setups. This is because of the wide number of variations that can exist. For rapid detection and tracking, the design of the face generates a certain amount of an issue. In the past few decades, numerous algorithms are implemented for face recognition and monitoring in a specified video stream [5]. The growing methodology does have perks and drawbacks. There will be some errors in any face tracking algorithm which will cause deviation from the necessary target. In this study, the tracker is one of the most effective approaches, that will be reliable and variability can be minimized. It is quicker and simpler, and we make use of the Eigenvectors for detecting the faces along with the superficial points of our face [6].

A Person's character is recognized with the aid of using his/her features reaction. The maximum critical factor is a facial reaction, used to differentiate someone from another. Each reaction has one-of-a-kind functions and features. So, face reputation performs an essential function in human behavior. In particular, a facial expression plays a main function inside the person to person communications and gives a very robust indication in assessing stages of a hobby of someone whilst interrelating with a device [12]. In this paper, the blind persons have a capable of perceiving human reaction because of face reputation and could get an audio message approximately the individual, the blind may be in a position to communicate to them without having to look

forward to the individual as opposed to returning to him and talking to him, only having to interpret the individual (provided the knowledge contained in the computer database is a person). The latest faces also can be brought to the record. The studies of expression reputation in PC subjects particularly specialize in the characteristic extraction and characteristic category. There are so-referred to characteristic extraction to extracting functions that may be used for category from entering photos or video streams [13].

There are many techniques for characteristic extraction. According to the form of facts enter; the prevailing techniques of characteristic extraction may be separated into two types: first is primarily based on stagnant pictures, and the alternative is lively pictures in series. Feature extraction techniques are primarily based totally on static pictures encompass Gabor wavelet transform, Haar wavelet transform, Local Binary Pattern (LBP), and Active Appearance Models (AAM) [9-11]. A rapid R-CNN (Faster Regions with Convolutional Neural Network Features) facial features popularity technique is proposed to put off the complex technique of expressive characteristic taking out and less-degree statistics strategy in the conventional facial features popularity.

2. RELATED WORK

Face detection is a PC generation that determines the region and length of the human face in an arbitrary (virtual) picture. The facial functions are detected and some other gadgets like trees, homes, our bodies, etc. It may appear as a selected case of object-elegance detection, in which the undertaking is locating the region and sizes of all gadgets in a picture that belongs to a given elegance [2]. Face detection can be seen as a trendier example of face position in face location; the task is to locate the positions and sizes of a known set of faces (commonly one). There are essentially two kinds of approaches for the detection of facial components within the given image, i.e. the basic feature and the level solution of the video. The feature-based approach tries to extract face features and balance it against facial feature information. While the image base strategy tries to get the best fit between preparation and image testing.

The active Shape Model is a center on a difficult non-rigid character such as definite Internal and higher-level presence of roles. This means that Active Shape Models (ASMs) are intended to identify the landmarks automatically that described the facial characteristics such as eyes, ears, nose, mouth, and eyebrows [14] in sequence. An ASM training method entails the compilation of a mathematical, facial model from a training set of objects that are landmarks systematically annotated. Karl Pearson has discovered the principal components analysis (PCA) which entails a mathematical way that transforms the probably correlated variables into some of the uncorrelated variables. Fundamental components are associated with unique variables through an orthogonal change. The process is described in such a way that the first primary element has as

large a heterogeneity as necessary (i.e. it reflects as much uncertainty as feasible in the information). And also, every successor factor does have the maximum possible deviation under the restriction it is exactly equivalent to the previous PCA which is immune to the comparative sizing of the independent variable [15]. Depending on the field of operation, it is often referred to as the discrete Karhunen – Loève transformation (KLT), or proper orthogonal decomposition (POD). The main advantage of PCA would be that it greatly reduces the dimension of the data and allows a test picture to be recognized. The trained images are not saved as image files but saved as their measurements, which are used to project any trained image to the set of the individual faces [16]. The appropriate data is must be collected and processed effectively, to measure within each picture. Similarly encoded database with one face encoding utilizing information theory terminology. The method behind takes out this type of data is to detain as numerous differences from the training images as possible.

Mathematically, the Eigen-face method is used to determine the most important additives of the face distribution. First, the covariance matrix eigenvectors of the collection of face pictures are determined after which they are looked at following their corresponding proper values. Then a threshold value of its Eigen is taken into account and eigenvectors with their Eigen-values are discarded much less than those threshold values. So the eigenvectors with the maximum great proper values are selected in the long run. Then the collection of face pictures is planned into the great eigenvectors to achieve a fixed one referred to as an individual faces. The great M Eigen-faces from an M dimensional subspace is referred to as "face space". Each character's face may be represented precisely because the linear aggregate of Eigen-faces can be approximately received the use of the maximum great Eigen-values.

Based on simple CNN architectures, pervasive item detection is performed with bounding box regression. This significant item detection is performed with improved neighborhood evaluation and pixel-stage segmentation [10]. Face detection and pedestrian detection are closely related to well-known object detection. Specifically, it performed with multi-scale adaptation and multi-characteristic boosting [21]. The dotted traces suggest that under positive conditions the corresponding domain names are related to every different object. It has to be noted that the domain names covered are diversified. Pedestrian and face images are common systems for well-known gadgets and image views, which have extra complicated versions of geometric systems and layouts. Exceptional deep fashions are therefore required for various geometric images.

3. PROPOSED SYSTEM

The proposed method utilizes the basic ideas of the KLT Algorithm and Viola-Jones Algorithm. The system explores the idea of finding the Eigenvectors for streaming video and

the quantity of extensive work has been achieved for real-time applications of video streaming.

Initially, the system senses the face using objects from the cascade object detector program to assess a facial location in a video sequence. The cascade object detector uses the Viola-Jones detection algorithm, and a qualified classification model. The KLT algorithm is used for calculating several feature points via the video sequence to classify facial features to verify them. When detecting the face, it recognizes facial feature points that could be accurately monitored. Tracking is used to point the locations with the specified facial points which use the visualization to map objects within the point tracking program. The point detector attempts to locate the coordinate system for each

point within the preceding frame in the consecutive frames. The approximate geometric transform function is then used to calculate the transformation, rotating, and scaling between both the new and the old points. The transformations of the new point are added to the boundary box around the face and finally monitoring the face.

In the proposed system has five steps to implement the concepts, they are getting the input Image, applying to pre-process on the Image, selecting the required points on the face, Extracting the Features, and applying CNN Classifier to recognize the Emotion.
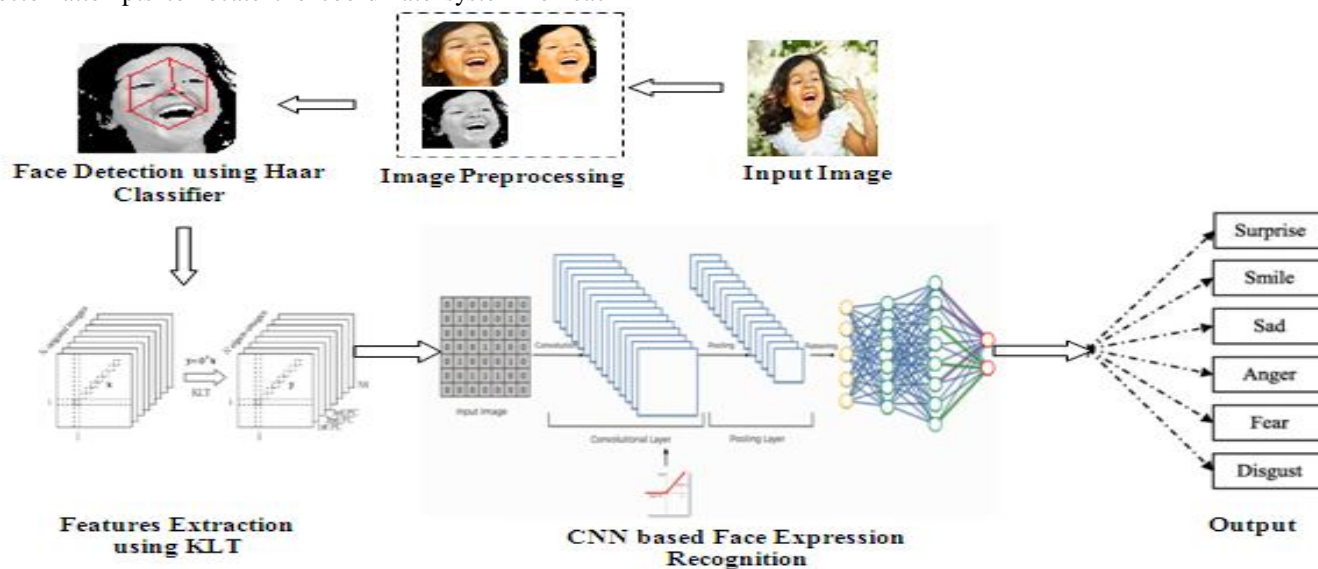

Figure 1: System Architecture

3.1 Input the Image

The first step is to give the input image of the human being. The image can be either colored or gray image files of a human face with the most-used image extension (jpeg, BMP, tiff, etc.). The object shall be of the moderate standard at least with the lowest resolution of 220 pixels in width, 220 pixels in height, and 92 pixels in minimum depth. Because the resolution is less than the minimum depth then a specified selection of the points would be very difficult. There are also certain constraints on the picture of the facial expression are required to preserve the specific traits and attributes that can be carried to the next level.

The most noticeable drawback was that the facial expression is frontal. The face needs to be in front and it does not turn at any angle and does not bend in any direction, and the features of the face are simple. This implies that the person does not wear a class or other mask to make sure the user can pick the points.

3.2 Facial Expression Pre-Processing

In pre-processing apply to achieve a good image with one standard in terms of contrast and size. This will help in the next steps used in the proposed approach especially in point

selection and getting the face features like contrast, size, and resolution

3.3 Face Detection

The face detection framework of Viola and Jones uses Haar basis feature filters. The shape of the individual face is ready to select the point needed by using 64 points distributed over the human face. And use these points for feature extraction. The choice of these points is to determine the shapes of each facial feature (eyes, eyebrows, and mouth) according to the point location that is not standardized. However, it depends on the features that will be extracted, and used for the classifier. Each point is determined by two axes X and Y, and after selecting each point in the image. We have to save the points for each image in a text file. The saved points can be loaded and shown in fig 2.
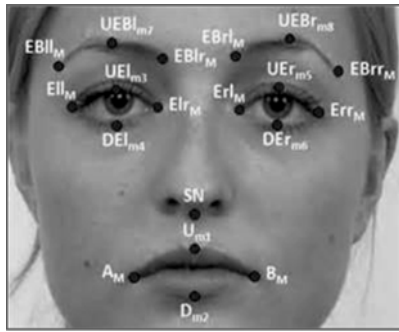
Figure 2: Feature detection

3.4 Features Extraction

Tracking a collection of characteristic points over the video frames by using the KLT algorithm, when the identification recognizes the features correctly, and then to identify the characteristic points which can be accurately monitored. In this stage calculate 28 features. The features are the distances between certain points explained in the previous stage. The shape of the face is often variable as the human emotion varies accordingly. This means that the properties of face components change when the facial expression is altered. The distances between the points are changed as a consequence of this. As will be seen later the distances between such points reflect the human emotion. The output of this stage is a certain length between certain points, and these lengths are the image characteristics.

3.5 Face Expression Recognition

CNN has gained considerable interest in addressing numerous dynamic classification issues. It was commonly used on facial expressions in face recognition and age prediction. We use of CNN, a deep classification according to the backpropagation algorithm for the classification task of the emotions. The derived characteristics are representing twenty-eight inputs and six outputs expressing the emotions, and also a dense node of sixteen points chosen while attempting to search for better outcomes.

Convolution Neural Network is a form of the deep neural network [20] which is most commonly used for image processing. These are also known as artificial neural networks or space invariant based on their architecture of mutual weight. These provide applications in video processing, recommendation method, image detection, medical image analysis, natural language processing, and economic power.
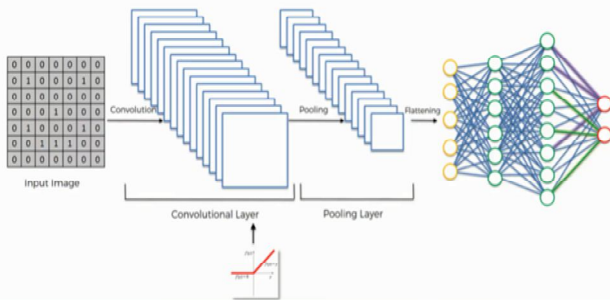

Figure 3: Convolution Neural Network

A Convolutional Neural Network could be an algorithmic deep Learning system is shown in fig. 3 that could absorb an input image, allocate importance (learnable weights and biases) to numerous aspects/objects within the image. In the pre-processing needed a ConvNet. The convNet classification algorithm is shown in fig 3. Whereas in the primitive way of filters are hand-engineered, with enough coaching. ConvNets have the flexibility to be told these filters/characteristics. Since the filtering is hand-designed in primitive methods, ConvNets has the opportunity to know with appropriate training these filters/characteristics. Max pooling is a combining process which extracts the expanded access from the Filter covered function map region. Therefore, the performance after the max-pooling layer will be a feature map that would contain the far more popular features of the prior convolution layer.

One approach is to aggregate metrics for these characteristics at different locations like c;alculating the max, min, or mean value of a region over an image. This operation is called the pooling operation, which is shown in fig 4. Therefore it is not mandatory to provide a pooling layer but it is advisable to decrease the rate of extracted features to prevent overfitting. The pooling primarily helps to remove sharp and smooth characteristics. Minimizing uncertainty as well as computational methods is often achieved. Max pooling requires the elimination of low-level characteristics like curves, loops. While average-pooling is really for streamlined features.
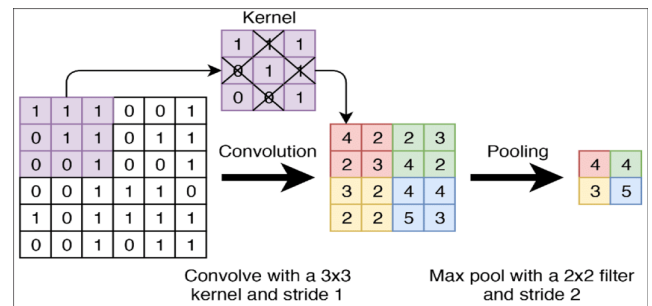

Figure 4: Pooling Layer

4. EXPERIMENT RESULT

Classification of facial expression involves vast quantities of data to represent the complexity of real-world circumstances. Public databases endorse research activities and provide an effective context for researchers' work. For this image classification, the Facial Expression Recognition data set is used. The file includes 35,887 images of faces with 48 * 48 pixels of grayscale. Seven types exist: Anger, Disgust, Fear, Good, Sadly, Shock, and Negative. The thing is to store these "files" in CSV format. In contrast to the tabular data stored in CSV format, the "columns" name is the first row: emotion, pixels, and consumption.

For pose variations, identification of faces depends on how robust the device is. If the system's goal is to only identify frontal faces, then use only a few classifiers and functions. For each face reaction, the number of objects depends on image

training and test sets. Identification from various angles depending on the image form and the training set appropriately, with at least one image per person for each pose. The quantity of pictures on the rest of the training set and assessment depends on the use of the program. The proposed method is to detect an individual from a collection of data that contains numerous images of an individual and different image of various individuals. Train the algorithm with a collection of images of the particular individual's face reaction and check on a collection of images containing various reacting images of the target person individually. This task will be a binary classification and exactness can be calculated with the aid of precision and recall. The confusion matrix and output of the experiment are shown in fig. 5 and fig. 6 respectively.

Accuracy, precision, and recall evaluation factor values for seven type's emotions we have taken, the resulting graph are shown in fig 7 and fig 8. Happy Emotion is the most commonly observed, as it has the most examples. Owing to enough cases, Sad, Shock, Negative, and Anger are also excellent in detection. Fear and disgust are doing worse, a potential cause for fewer examples of training and disgust is very similar to rage features. Sad feelings are often similarly identified as neutral, with only that much data it is difficult to differentiate them.
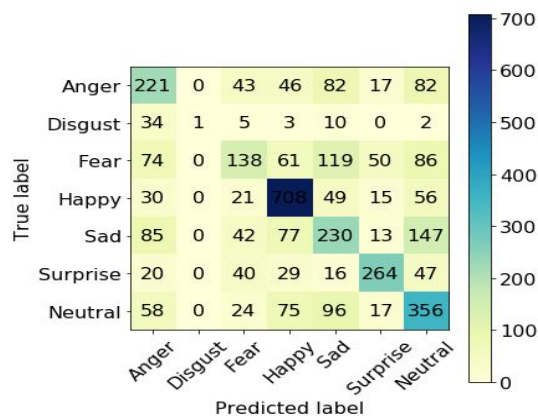

Figure 5: Confusion Matrix

> testing image index: 1607
> true emotion: happy
> predicted emotion: happy
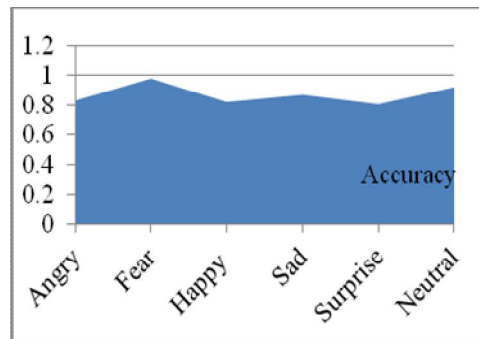
Figure 6: Output of Experiment
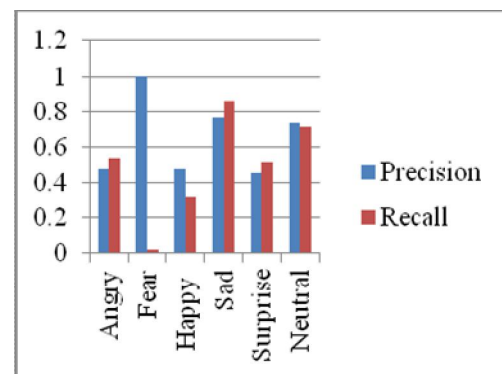

Figure 7: Accuracy

Figure 8: Precision and Recall

5. CONCLUSION

The computational models were implemented in this paper and chosen after extensive research, and effective test results, indicate the researcher's option was accurate. The system with face emotion detection and automatic face recognition did not have recognition precise over 90%, despite the small number of Eigenfaces used in the PCA transform. In this experimental research, this mechanism was evaluated beneath very stable conditions, and real-world performance visualized to be much more reliable. The completely automated facial images recognition system showed almost perfect accuracy and there is no need to do any more work in this field, the researcher believes. The solution mentioned would automatically acquire the model features and the insufficiency caused by artificial structural elements. The new framework by learning analyzed data information directly extracts the pixel information of the image. The suggested methodology could to some degree increase the classification accuracy of facial features in cluttered images. The proposed framework also achieves a better acceptance rate.

REFERENCES

1. T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui. **MPED: A multi-modal physiological emotion database for discrete emotion recognition**, *IEEE Access*, Vol. 7, pp. 12177-12191, 2019.
2. Ming Hsuan Yang, J. David Kriegman, and Narendra Ahuja. **Detecting Faces in Images: A Survey**, *IEEE Trans. PAMI,* Vol. 24, No.1, pp. 1-25, 2002.

3. M. Jones, and P. Viola. **Rapid Object Detection using a Boosted Cascade of Simple Features**, *Proc CVPR IEEE.* pp.511-518, 2001.

4. S. P. Himani, G. T. Darshak, and K. Jal Udesang. **A Survey on Object Detection and Tracking Methods**, *Int J Innov Res Comp Comm Engg*, Vol. 2(2), pp.6 -10, 2014.

5. M. H. Yang, N. Ahuja, and D. Kriegman. **Face detection using mixtures of linear subspaces**, *Fourth IEEE Int Conf Autom*, Vol. 24(1), pp: 70-76, 2000.

6. C. L. Huang, and C. W. Chen. **Human facial feature extraction for face interpretation and recognition**, Pattern Recog. Vol. 25, pp. 1435–1444, 1992.

7. L. Yuille, P. W. Hallinan, and D. S. Cohen. **Feature extraction from faces using deformable templates**, *Int. J. Comput. Vision 8*, pp. 99– 111, 1992.

8. H. Meng, N. Bianchi Berthouze, Y. Deng, J. Cheng, and J. P. Cosmas. **Time-delay neural network for continuous emotional dimension prediction from facial expression sequences**, *IEEE Trans. Cybern.,* Vol. 46, no. 4, pp. 916-929, 2016.

9. B. Yang, X. Xiang, D. Xu, X. Wang, and X. Yang. **3D palmprint recognition using shape index representation and fragile bits**, *Multimedia Tools Appl.,* vol. 76, no. 14, pp. 15357-15375, 2017.

10. J. Cai, Q. Chang, X. L. Tang, C. Xue, and C. Wei. **Facial expression recognition method based on sparse batch normalization CNN**, *Proc. 37th Chin. Control Conf. (CCC)*, pp. 9608-9613, 2018.

11. N. Kumar, and D. Bhargava. **A scheme of features fusion for facial expression analysis: A facial action recognition**, *J. Statist. Manage. Syst.,* Vol. 20, no. 4, pp. 693-701, 2017.

12. S. H. Lee, K. N. K. Plataniotis, and Y. M. Ro. **Intra-class variation reduction using training expression images for sparse representation based facial expression recognition**, *IEEE Trans. Affect. Comput.,* vol. 5, no. 3, pp. 340-351, 2014.

13. S. K. A. Kamarol, M. H. Jaward, J. Kalviainen, H. Parkkinen, and R. Parthiban. **Joint facial expression recognition and intensity estimation based on weighted votes of image sequences**, *Pattern Recognit. Lett.,* Vol. 92, pp. 25-32, 2017.

14. D. O. Medley, C. Santiago, and J. C. Nascimento. **Deep Active Shape Model for Robust Object Fitting**, *IEEE T Image Process,* Vol. 29, pp. 2380-2394, 2020.

15. N. H. Barnouti, Mayyahi, and S. Dabbagh. **Real-Time Face Tracking and Recognition System Using Kanade-Lucas-Tomasi and Two-Dimensional Principal Component Analysis**, *Int. Conf. AdvEngg, Duhok,* pp.24-29, 2018.

16. R. Kaur, and E. Himanshi. **Face recognition using Principal Component Analysis**, *IEEE Int Adv Comput (IACC), Banglore,* pp. 585-58, 2015.

17. M. Ramalingam, and R. M .S. Parvathi. **Secure Semantic Aware Middleware: a Security-Based Semantic Access Control for Web Services**, *International Review on Computers and Software,* Vol. 8, No. 9, pp. 2136-2141, 2014.

18. M. Ramalingam, and R. M .S. Parvathi. **Policy-Based Semantic Access Control Framework for Fine-Grained Access in Semantic Web Services**, *European Journal of Scientific Research,* Vol.74 No.1, pp.154-163, 2012.

19. D. Saranya, S. Thulasidass, and D. Gomathi. **Automatic Service Discovery using Ontology Learning Semantic Focused Crawler for Mining**, *International Journal of Advanced Research in Computer Engineering & Technology.* Vol. 4, No. 10, 2015.

20. Sheetal S. Pandya, Nilesh B. Kalani. **Review on Text Sequence Processing with use of different Deep Neural Network Model**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 8, No.5, pp. 2224- 2230, September - October 2019.

21. P. M. Ashok Kumar, Karishma Begum Shaik, Divya Teja Ravuri, **Multi-view Face Recognition Using Deep Learning**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No.3, pp. 2224- 2230, May - June 2020.