

## Face Tracking using Convolutional Neural Networks

Dr.K.Kiran Kumar<sup>1</sup>, Dr.Arvind Yadav<sup>2</sup>, Dr.B.B.V.Satya Vara Prasad<sup>3</sup>,  
Dr. M.Ramesh Kumar<sup>4</sup>

<sup>1,2</sup>Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,  
Vaddeswaram, AP, India.

<sup>3,4</sup>Department of Electronics and Computer Engineering, Koneru Lakshmaiah Education Foundation,  
Vaddeswaram, AP, India.

### ABSTRACT

Face detection and recognition is very active re- search area in the field of computer vision. It is one of the crucial and preprocessing steps for most of the Computer vision based application especially where human interaction is necessary. Face detection and recognition is done in many approaches like feature extraction and classification, Statistics based and image based etc. However, it is still complex problem because, of several reasons. In this work, we try to solve above mentioned problem using feature extraction and classification methods. The experiments are conducted and performance is evaluated over the various databases like Yale and AR. The Face detection and recognition problem is also solved using convolution neural networks. Results are validated on videos related to Indian movies.

**Key words:** Face detection, recognition, feature extraction, CNN.

### 1.INTRODUCTION

Computer vision is the science and technology of machines that see, which means that machine extracts information from an image necessary to solve a particular task. There are number of applications for computer vision like surveillance, criminal identification, video conferencing and intelligent human computer interfaces etc. To apply most of these applications to real world, the basic preprocessing stage is to detect the face in a video and recognize that particular face.

The face detection and face recognition is one of the long standing and bio metrics measure problem. Most of the researchers contributed their efforts to solve since from 2 decades back [1]. The recent developments of face detection technology is already implemented in modern cameras to capture photos efficiently. Face recognition is implemented in popular social networking sites and in modern laptops for authentication purposes. Surprisingly there are still some issues with face detection and no unique technique that works well with all types of applications because of constraints like different lighting conditions, scales, rotations, perspective variations, face patterns and poses. The problem of face detection and face recognition chosen to solve in this paper looks like simple and older but there are some interesting facts that are enforced to choose this

particular problem. For example a method which has obtained 98% accuracy on particular faces data set giving only 50% on another face dataset. Through this motivation, problem of face detection and face recognition is chosen.



**Figure 1:** Sample Images from Yale (first row) data set and AR (second row) fair, eye closed and bad lightning respectively

To implement the present problem of research there are three important building blocks. First is feature extraction second is classifier, and third is on what datasets solution is performing. PCA-principal component analysis, LDA-linear discriminate analysis, Dsift, and LBP-Local binary patterns are chosen to extract features. The second building block is classifier, KNN- k nearest neighborhood, SRC classifiers are used in this work. In addition the recent classifier called Convolution Neural Networks CNN is also used for classification in videos. Finally dataset Large numbers of face datasets that are available to researchers, Yale and AR are considered for the present problem. In addition IMFDB data set on Indian movies is also used for results analysis.

The rest of the paper is organized as follows: Related work is explained as Section 2. Original problem of face detection and face recognition with the flow of implementation is described in Section 3. This section comprises a comparative study of the solving the problem on various data sets like Yale [2], AR [3] with various classifiers and feature extraction methods. Experimental results that are obtained after applying the face detection and face recognition on Indian movies with analyzed results and CNN are discussed in Section 4. Finally conclusion and future scope of the problem is given in Section5

**2. RELATED WORK**

As discussed in introduction section three building blocks of the present problem are face detection, face recognition and face data set. This section describes the related work corresponds to each building block respectively.

**A. FACE DETECTION**

Face detection is an important and fundamental problem in computer vision and there have been many attempts to address it [4]. In early 1970s, heuristic techniques were applied [5]. These methods are restricted only to plain background, frontal face and typical passport photograph scenario. Over the past decade, there has been a great deal of research interest spanning several important aspects of face detection.

Since face detection techniques requires a priory information of the face, they can be effectively organized into two broad categories distinguished by their different approach to utilizing face knowledge. The techniques in this category make explicit use of face knowledge and follow the classical detection methodology in which low level features are derived prior to knowledge-based analysis [6]. The apparent properties of the face such as skin color and face geometry are exploited at different system levels. Typically, in these techniques face detection tasks are accomplished by manipulating distance, angles, and area measurements of the visual features derived from the scene. Since features are the main ingredients, these techniques are termed the feature-based approach. The other important and successful milestone in face recognition is viola- Jones algorithm[7].

Image-based representations of faces [8] are directly classified into a face group using training algorithms without feature derivation and analysis. Unlike the feature-based approach, these relatively new techniques incorporate face knowledge implicitly [9] into the system through mapping and training schemes. A new hierarchical face detection framework is proposed [10] basing on grayscale behavior of faces in pyramid images. When the resolution of a face image is reduced gradually either by sub sampling or averaging, pyramid features of the face will disappear. At low resolution, face region will become uniform. Face detection using feature based approach is more efficient than other methods because it is invariant to transformation and size.

**B. FACE RECOGNITION**

A formal method of classifying faces that is Face recognition was first proposed by Francis Galton in 1888 [11]. The important mile stones in the development of face recognition can be classified in to segmentation based, Feature extraction based, Methods based on Statistics approach and some unsupervised methods using neural network. Segmentation or classification based approaches mainly concentrates on learning. These approaches has changed the face recognition scenario from single face to the multiple faces. Most of the researchers focused in this area after 1990. There are number of mathematical models proposed and implemented successfully basing on the learning process [12]. The PCA has been intensively

exploited in face recognition applications. Many other linear projection methods have been studied too. The LDA (Linear Discriminate Analysis) [13] has been proposed as a better alternative to the PCA. Neural networks methods have also been applied to face recognition. The successful examples are Probabilistic Decision Based Neural Network and Evolution Pursuit (EP) methods. The main advancement in Statistics approach is by Eigen faces. Eigen face recognition was first proposed by Sirovich and Kirby [14]. It is considered the first successful example of facial recognition technology. Eigen faces approach coupled with Hausdorff distances is also proposed for face recognition. Hausdorff distance measures are incorporated with the eigen-weighting function so that distances at regions of important facial features will be emphasized. Methods based on Feature extraction and classification are more efficient compared with all the remaining. The present research also focuses on same.

**C. FACE DATASET DESCRIPTION**

Face datasets can be categorized based on two settings, constrained and unconstrained. In constrained setting, Yale, AR, CMU, etc are the popular databases. For unconstrained setting, LFW, Pub fig databases are used. The table 1 describes the face dataset description used in this work. Also figure-2 describes the sample images from IMFDB dataset

**Table 1:**Face dataset description

Dataset name	Description
Yale	It contains frontal grayscale face images of 15 people, with 11 face images of each subject, giving a total of 165 images. Lighting variations include left-light, center-light, and right-light. Spectacle variations include with-glasses and without-glasses. Facial expression variations include normal, happy, sad, sleepy, surprised, and wink.
AR	4,000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf).
IMFDB	This face database consisting of 34512 images of 100 Indian actors collected from more than 100 videos. Dataset has high degree of variability in terms of scale, pose, expression, illumination, age, resolution, occlusion, and makeup. Data set is provides with annotation of every image in terms of age, pose, gender, expression and type of occlusion that may help other face related applications.



**Figure 2:** Sample Images from IMFDB data set with wide variety of ariations

#### D. DESCRIPTION OF FEATURE EXTRACTION AND CLASSIFICATION

The major parts of current work are Feature extraction and classification. Any given image can be decomposed into several features. Feature is any distinctive aspect, quality or characteristics such as example color, height etc. The combination of  $n$  features is represented as a  $n$ -dimensional column vector called a feature vector. Then-dimensional space defined by the feature vector is called feature space. The main objective of a feature extraction technique is to accurately retrieve features. A common way to resolve this problem is to use dimensionality reduction techniques. Two of the most popular techniques for this purpose are: Principal Components Analysis (PCA) and Linear discriminate Analysis (LDA). In addition to PCA and LDA the dsift and LBP were also can be used.

For the PCA algorithm Input is an  $m \times n$  matrix, where  $m$  is the number of parameters and  $n$  is the number of samples. First step is Computation of covariance matrix of the dimensions and Find eigenvectors of covariance matrix. Next step is to sort eigenvectors in decreasing order of eigen values. Finally Project onto eigenvectors. LDA is another supervised dimensionality reduction technique. It projects high dimensional data onto a lower dimensional space by maximizing the separation of data points from different classes and minimizing the dispersion of data from the same class simultaneously, thus achieving maximum class discrimination in the dimensionality-reduced space.

The classifier uses the feature vector [15] normalized to assign the input data to the different classes. Usually any type of data is divided in to training and test sets. Classification is a two step process. Learning step uses training set and learn model. Classification uses test set and estimates the accuracy of classification. KNN classifier [20] is a supervised learning system, which has to be trained with data to correctly set the involved internal parameters. Though other methods based on neural networks or support vector machines could have been selected, we chose this model for simplicity and because it has been widely used to address similar pattern recognition problems. The KNN and SRC are used for classification.

KNN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. In the model, called  $k$ -nearest neighbor (kNN), we use two stages. In the first stage, if we find that a given class has more training samples closer to the test pattern, then we declare this class as an outright winner and allocate the test pattern to this class. However, if we find there are an equal number of highest neighbors for more than two classes surrounding the test pattern, then we perform the second stage called conflict resolution. At this stage, the class whose distance from test data averaged over all its training samples within the hyper sphere is found to be the smallest is declared the winner.

SRC(Sparse Representation Classier) is also implemented for the present analysis. Generally exploiting sparsity is critical for the high performance in the classification of high dimensional data such as face images. With sparsity[16] properly harnessed, the choice of features becomes less important than the number of features used. Moreover,

occlusion and corruption can be handled uniformly and robustly within the same classification framework. It can achieve as triking recognition performance for severely occluded or corrupted images by a simple engineering representing using SRC algorithm with no special.

Local Binary Pattern (LBP) descriptor is an efficient facial features extraction method; it extracts information from neighboring pixel values and develops histogram of the image[17]. This is a non parametric operator and describes the local spatial structure of an image. In detection or identification of face scale is very important that is how the face is small or big the algorithm should work efficiently. Scale Invariant Feature Transform (Dsift) addresses this problem using image pyramids [18].

The experimental result of the present problem divided into two parts. First part is analysis of KNN classifier by changing dimension over the methods like PCA, LDA, Dsift and LBP as feature extraction. The same experiment is conducted in the case of Yale and AR databases. Table 2, and Table 3 represent respective experimental values.

**Table 2:** Accuracy of Feature extraction over Yale data set

Feature extraction	K=1	K=3	K=5	K=7
PCA	60.9959	58.2573	58.3402	55.3527
LDA	84.1494	83.3195	85.3942	82.9046
Dsift	97.1784	97.5104	97.5934	98.0913
LBP	75.8506	75.4357	75.3527	75.6017

**Table 3:** Accuracy of feature extraction over AR data set

Feature extraction	K=1	K=3	K=5	K=7
PCA	63.4670	59.7421	54.8711	47.5645
LDA	15.3295	15.7593	16.6189	14.7564
Dsift	88.2521	88.8252	89.8281	85.5301
LBP	80.5158	78.9398	75.9312	70.3438

Feature extraction methods also studied over two datasets specified. Table 4, and Table 5 represent respective experimental values.

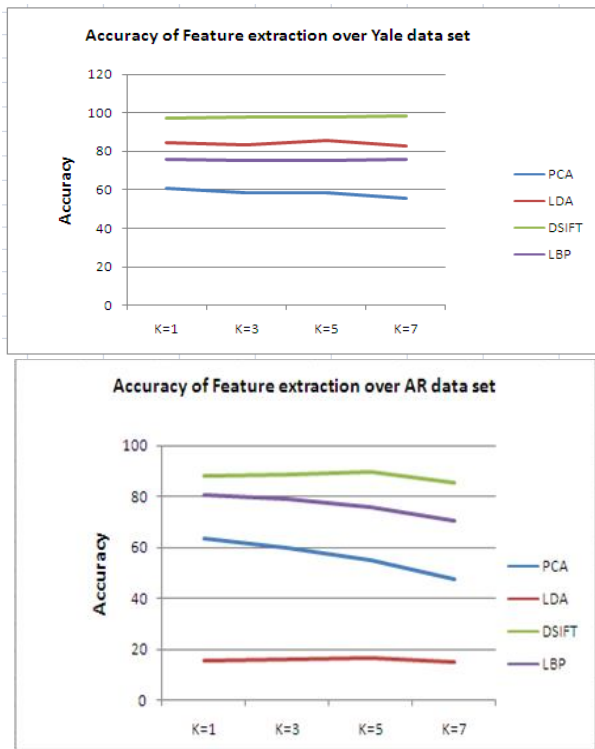
**Table 4 :**Accuracy of feature extraction over data sets using KNN classifier

Dataset	PCA	LDA	Dsift	LBP
Yale	60.9959	84.1494	97.1784	75.8506
AR	63.4670	45.3295	88.2521	80.5158

**Table 5:** Accuracy of feature extraction over data sets using SRC classifier

Dataset	PCA	LDA		Dsift	LBP
Yale	87.8008	79.7510		99.6680	96.7635
AR	83.9542	41.4613		93.9828	95.4155

The variation in feature dimension does not have much impact on Yale and AR datasets performance. These are shown in figure-3 and 4 respectively. Classification algorithms typically employ two phases of processing one is training and another is testing. In the initial training phase, characteristic properties of typical image features are isolated and, based on these, a unique description of each classification is created [19].



**Figure 4:** Accuracy of feature extraction over AR dataset

**E. DESCRIPTION OF CONOLUTIONAL NEURAL NETWORKS FOR CLASSIFICATION**

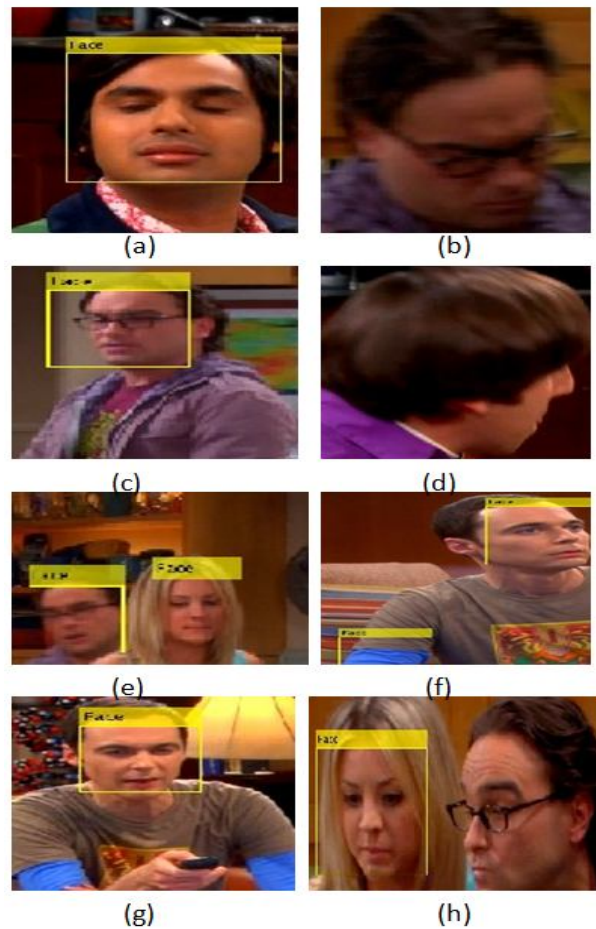
First face detection is performed in this work. A bounding box obtained on face denotes that face is detected. The figure 5 shows the face detection results in different cases. Results are obtained on online video clips.

**Input:** A set of images of the person

**Algorithm:**

1. Input image sent to convolution layer to extract features from image using filters.
2. A feature map is extracted which is the output of convolution layer and input of pooling layer.
3. Pooling layer reduces the dimensionality of the features extracted but the original data remains same. Max-pooling technique is implemented.
4. Next we flatten the matrix data to vector form and feed it to the fully connected layer and combine the features together and then a model is generated.
5. Finally, we have a SoftMax layer which helps to classify the output.

**Output:** A recognized face image of the person



**Figure 5** Analysis of face detection in different cases (a) Face is detected in front view (close) (b) Face is not detected due to lightning (c) Face detected in side view (d) face is not detected due to pose variation (e) Multiple Face detection (f) Face and hand detected as face (g) Face is detected in front view (h) Detected and undetected faces

This section also describes the Convolution Neural Networks (CNN) for classification. The framework is shown in figure-6. The algorithm describes the detailed steps in classification process.

Architecture mainly consists of two sets of convolution and pooling layers. After these layers features are extracted and flatten vector is created which can be given as input to fully converted layers. These layers can perform actual classification process. Here basic unit is node. Two operations called pre-activation and activation can be done in each node. In the final step softmax activation function is used to obtain probability scores of the classes.

CNN model is applied for the three datasets. The classification accuracies of 95.4%, 94% and 89.9% are obtained for three datasets Yale, AR and IMFDB [20] respectively using CNN. Figures 7 and 8 shows results obtained on videos using IMFDB.

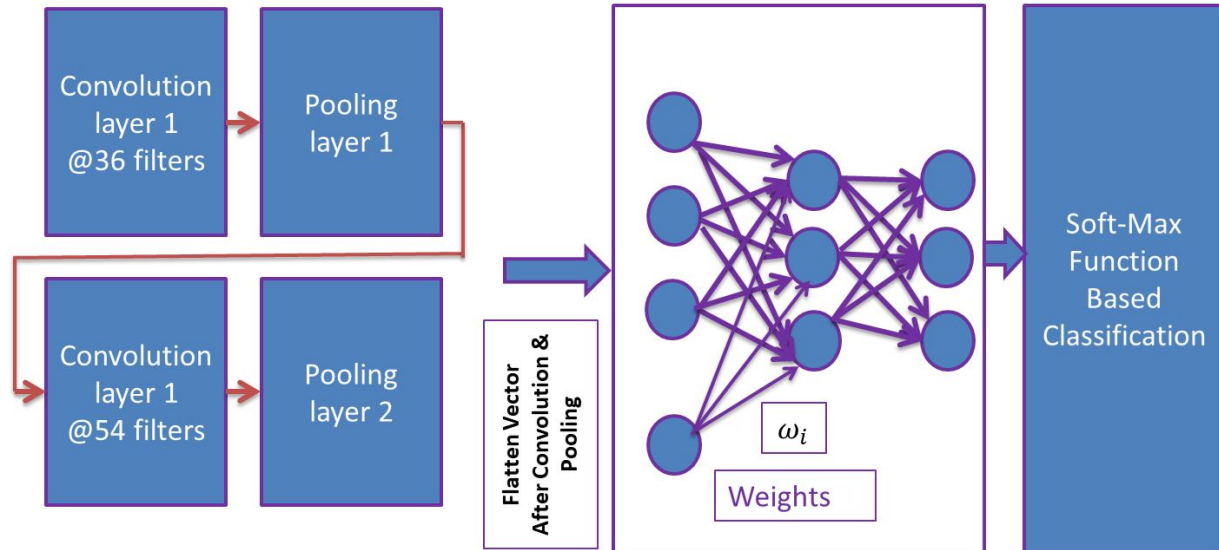


Figure 6: CNN Framework for the Classification model

### 3.CONCLUSION

In this paper, we tried to solve the problem of face detection and recognition. KNN and SRC based classifiers are used to investigate the robustness of PCA, LDA, SIFT and LBP features for various images on face recognition. We carried out experiments on Yale and AR datasets. The experimental results shows that the LDA is algorithm for feature extraction is superior to the PCA. For the case of classifiers KNN is some what in efficient than SRC classifier. but KNN is very popular, applicable to various situations due to its siplicity. The algorithm called CNN gives best results compared to KNN and SRC. Future work includes implementation of algorithm and test for changes of illumination, face pose, facial expression etc. The results obtained in the present problem are successfully extended to Indian movie database called IMFDB.

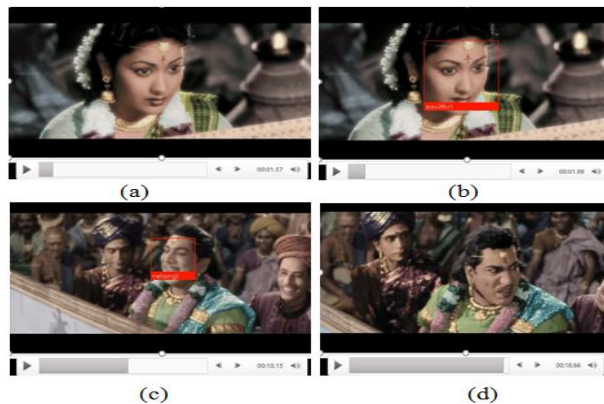


Figure 7: Face detection and recognition in videos sample movie clip-1(a) frame-1 (b) frame-2 (c) frame-3 (d) frame-4



Figure 8: Face detection and recognition in videos sample movie clip-2(a) frame-1 (b) frame-2 (c) frame-3 (d) frame-4

### REFERENCES

1. R. C. W. ZHAO, "Face recognition: A literature survey," *ACM Computing Surveys*, 2003
2. "Yale Face Database". "Online", 2006
3. R.B.A.M.Martinez, "TheARFaceDatabase," 1998
4. B. K. Erik Hjelmas, "Low face detection: A survey, *Computer Vision and Image Understanding*," 2001
5. T. K. T. Sakai, M. Nagao, "Computer analysis and classification of photographs of human faces," *First USA-Japan Computer Conference*, 1972.
6. P. Jahnavi<sup>1</sup>, Dr. E. Vamsidhar<sup>2</sup>, Dr. C. Karthikeyan<sup>3</sup>(2019), "Facial expression detection of all emotions and face recognition system", *International Journal of Emerging Trends in*

- Engineering Research, 7(12),PP.778-783, ISSN 2347 – 3983.
7. Dr Nishad Nawaz (2019), “**Artificial Intelligence Face Recognition for applicant tracking system**”, International Journal of Emerging Trends in Engineering Research, 7(12),PP.778-783, ISSN 2347 – 3983.
  8. Reddy A.V.N., Phani Krishna C. (2018), “**A survey on applications and performance of deep convolution neural network architecture for image segmentation disease classification from MRI images**”, Journal of Advanced Research in Dynamical and Control Systems, 10 (0), PP. 609-617
  9. Satyanarayana P., Jaya Devi N., Sri Hasitha S.K., Sessa Sai M. ( 2018) , “**An enhanced viola-jones face detection method with skin mapping & segmentation**”, Advances in Intelligent Systems and Computing, 668 (), PP. 485- 493
  10. Sucharitha G., Senapati R.K. ( 2018) , “**Local extreme edge binary patterns for face recognition and image retrieval**”, Journal of Advanced Research in Dynamical and Control Systems, 10 (), PP. 644- 654
  11. Jahnavi P., Vamsidhar E., Karthikeyan C. (2019), “**Facial expression detection of all emotions and face recognition system**”, International Journal of Emerging Trends in Engineering Research, 7(12), PP.778-783.
  12. Vishal P., Snigdha L.K., Bano S. (2019), “**An efficient face recognition system using local binary pattern**”, International Journal of Recent Technology and Engineering, 8(0), PP.912-914.
  13. Gopi G.S., Priyanka P., Bharathi K., Rajasekhar J., Kommuri K. (2019), “**Access control of door using face recognition and home security alert using raspberry pi and internet**”, International Journal of Innovative Technology and Exploring Engineering, 8(6), PP.1841-1845.
  14. Reddy B.K., Nikhil C., Kolluri M., Ramya K.R., Mandhala V.N. (2019), “**Securing cloud data using face recognition and deep learning**”, International Journal of Engineering and Advanced Technology, 8(4), PP.1857-1863
  15. Vijaya Prasad K., Kishore P.V.V., Srinivasa Rao O. (2019), “**Skeleton based view invariant human action recognition using convolutional neural networks**”, International Journal of Recent Technology and Engineering, 8(2), PP.4860-4867.
  16. Ahmadsaidulu S., Ravi Chandra D., Vinith U. (2019), “**Home security surveillance using face detection technique**”, International Journal of Innovative Technology and Exploring Engineering, 8(6), PP.1474-1477
  17. Bhargavi, V. Ratna; Rajesh, V(2018), “**Computer Aided Bright Lesion Classification in Fundus Image Based on Feature Extraction**”, INTERNATIONAL JOURNAL OF PATTERN RECOGNITION AND ARTIFICIAL INTELLIGENCE, 32(11). DOI:10.1142/S0218001418500349
  18. Ravi, Sunitha; Suman, Maloji; Kishore, P. V. V.; Kumar, Kiran E.; Kumar, Teja Kiran M.; Kumar, Anil D.(2019), “**Multi modal spatio temporal co-trained CNNs with single modal testing on RGB-D based sign language gesture recognition**”, JOURNAL OF COMPUTER LANGUAGES, 52, PP.88-102. DOI: 10.1016/j.cola.2019.04.002
  19. Kumar, EK; Kishore, PVV; Kumar, MTK; Kumar, DA, “**3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2-stream CNN**”, EURO COMPUTING, 2020. DOI:10.1016/j.neucom.2019.09.059
  20. Ahammad, SH; Rajesh, V; Rahman, MZU, “**Fast and Accurate Feature Extraction-Based Segmentation Framework for Spinal Cord Injury Severity Classification**”, IEEE ACCESS, 2019. DOI:10.1109/ACCESS.2019.2909583.