



Facial expression detection of all emotions and face recognition system

P. Jahnavi¹, Dr. E. Vamsidhar², Dr. C. Karthikeyan³

¹ A student in the Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India, pitchukajahnavi3@gmail.com

^{2,3} Assoc. Professor in the Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India, enireddy.vamsidhar@gmail.com, ckarthik2k@gmail.com

ABSTRACT

Nowadays we have seen that the study of facial expression recognition became most important. In humans, the way of interaction will possible through text and voice. Apart from that, facial expressions play an important role in smooth communication among individuals. We can recognize the emotion of a human, based on facial expressions. First of all, facial areas were removed from the original images. Meanwhile, by using the different types of optimizers we have trained the convolutional neural network mainly focused on the face extraction. The accuracy of other methods like 2D PCA is 78.61%, Adaboost is 83.12%, SVM is 87.15%. In this paper, the method we proposed is softmax classifier by which we obtained an accuracy of 94.82% which is the highest among all. Here this softmax classifier we have used in this experiment, suggests the combination of face detection and emotion detection using CNN.

Key words : CNN, Face detection, Emotion Detection

1. INTRODUCTION

Image processing is an important research topic and it has attracted many researchers around the world to work on it. CBIR content based image retrieval is a part of the Image processing in which input is an image and output is also an image. The Image pre-processing. Feature extraction, feature selection and classification is a part of the process. Classification of the images are done using different Data Mining techniques [13, 16, 17, 18], Neural networks, RNN, and also optimization of the networks are also done in order to improve the accuracy of the classification of Images[14,15]. With the increase use of machine learning and it is used in the different applications and KNN, Random Forest are some algorithms used in various applications[19,20]. There is a large increase in the technology that has improved the classification of images and also the recognition of images is a part in it. In nowadays there is a large need to detect the

faces and the emotions. We can recognize the emotion of a human-based on the facial expressions. A human can be fake through speech or text, but the inner mind can be expressed in the face [1] so nowadays the researchers are focusing on recognizing different types of smiles after extracting the features of a human face as it is the most important expression while communicating.[2]. This type of research has been used in many applications such as designing robots that are emotional, promoting psychology studies, predicting psychiatric patients ' social behavior, surveying customer satisfaction, and automated focusing.[3]. Though by doing the research in recognizing the smile after extracting the features, the expression of face is to be detected and it is not a simple task. we will face some challenges in extracting the features as there is no particular position of the head, lighting, the angle we are facing towards the camera, hair setting. So by these challenges, we will find it difficult to detect the face and to extract the features and to compare them to the training dataset.[1]. In 2012, to classify the smile, Guo and Bai had approached a method in which from the images the Pyramid Histogram of Orientation Gradient features are extracted first, later the linear discriminating analyzes are used to reduce measurements and finally the Adaboost and support vector machines have been used for the classification. [4]. The most commonly and frequently used method which deals directly with the entire image is the Support vector machine. We mostly care about the human face and the context which is present so this leads to the negative effects at the end of the results. In order to overcome these effects, the facial features are extracted first from the images and later implemented the method which is popular in deep learning which is CNN in order to construct the classifiers. [5]. In this paper by using the softmax classifier we recognized all the 7 emotions in a human face (happy, sad, disgust, anger, fear, surprise,neutral) . We will recognize these emotions when the web cam captures the image and by extracting the features of the image [5]. After comparing the features with the training dataset, the emotion is recognized.

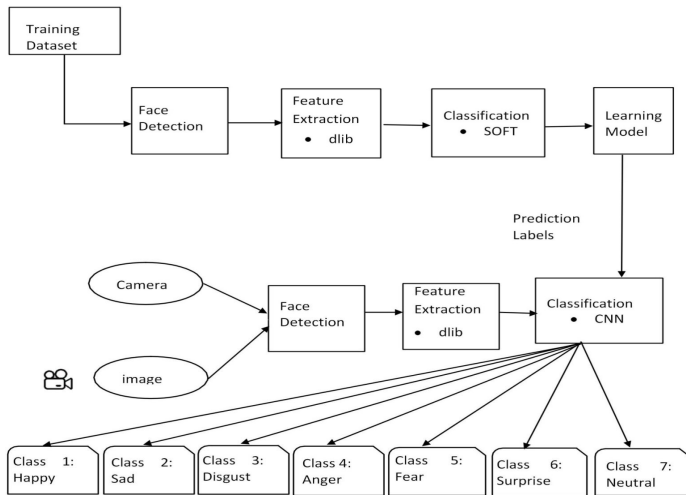


Figure 1 : Block diagram of the method

2. PROPOSED METHOD

In the experiment that is performed here have used two different datasets COHN-KANADE AU Coded Facial Expression Database which consists of 25,000 images. A total of 6481 images are used for training purpose and 1619 are used to test the network. The standard 48 X 48 pixel images are used here.

2.1 Cohn-Kanade Dataset

There are 100 university students in published portion of COHN-KANADE AU-Coded Facial Expression Database. We aged between 18 and 30 years of age. 65% are female, African-American was 15 percent, and Asian or Latino was three percent. An experimenter directed subjects that are used to perform twenty-three facial presentations which are in the series form that includes both single and multiple units of action. The images which are in sequence starting from neutral to the final target display were in the form of 480 to 490-pixel arrays in 640 which is having an accuracy of 8-bit grayscale. In the files of images, they contain other files called “Sequence” which are in the form of small text files that identify the order of the images that should be read to extract the features. There are 7 different types of emotions available which are Angry, Sad, Disgust, Fear, Surprise, Happy and Neutral.

Table 1 : Table showing the details of the COHN-KANADE database

Database	Sample Details	Available Descriptions
COHN-KANADE Database (also known as CMU-Pittsburg data-base)[6]	585 image sequences from 97 subjects. Age : 18 to 30 years Gender : 65% female Ethnicity : 15% African – Americans and 3% of Asians and Latinos.	“ Annotation of FACS Action Units and emotion specified expressions”

2.2 Facial expression recognition challenge (FER) Dataset

This collection contains a total of 213 images. For each object we have 10 different subjects and seven different emotions. For each emotion we have 20 images approximately. The seven emotions are Angry, Happy, Sad, Fear, Disgust, Surprise, Neutral.

2.3 Dataset Preparation

Two datasets were used to train and test the proposed system, namely Facial Expression Detection Challenge and COHN-KANADE. The Facial Expression Detection Data Set consists of 25,000 image sequences from 100 subjects, while the Facial Expression Detection Data Set consists of 7,000 images. We used 6481 images for learning and 1619 images for analysis from various Cohn-Kanade dataset samples for our study. Likewise, 6481 images were used for training, and 1619 images were used to check the facial expression recognition challenge dataset. Based on the facial image structure of 48* 48 pixels, initial images were taken. Automatic face detection was performed to identify the facial image by using our own system's face detector based on the haar classifier.

Table 2 : Table showing the details of the training, testing and validation number of Images

Dataset	Neu- tral	Ang- ry	Disgu- st	Fea- r	Hap- py	Sa- d	Surpri- se
Training Set	50	50	55	48	65	36	55
Validati- on set	31	29	35	25	33	23	36
Test Set	30	27	31	25	32	19	33

2.4 Face recognition

Besides, we can recognize and classify the smile and it is only concerned with the features of the human face. We can tell if a person is smiling or not by the conditions of eyes and mouth. The presence of the context can damage the classifier’s predictive accuracy. The facial areas are first removed from the original images to prevent the noise from the background. There are two types of face detection approaches that are commonly used. The first approach is based on the color of the skin we will recognize the face. The main idea of this approach is identifying the face region in the color space based on skin colour[7]. And the second approach which is adopted in this paper [8] is introduced in 2001 by Paul Viola which is a combination of features of haar – like and the Adaboost which is used in face recognition.

2.5 CNN

This is a network which contain many layers between the input and the output layers and also known as deep Learning algorithm which takes an image as an input, assign significance to dissimilar aspects/objects in an image, and can differentiate one from the other [9]. In a ConvNet, pre-

processing is much smaller than other classification algorithms. ConvNets has the ability to learn these characteristics/filters.

A ConvNet's structure is similar to the communication of the human brain which contains the neuron patterns and this was influenced by the Visual Cortex's Organization. The Receptive field which is the only restricted area of the visual field, the neurons which are individual in that receptive field respond to stimuli. In order to cover the whole visual region, such fields are overlapped in the form of sets. Because of the reduction in the number of parameters involved and the reusability of weights, For the image dataset, a better match is provided by the architecture. In other words, the network can be trained. Yes, It is a trainable architecture that is biologically based and the features which are invariant can be learned[10]. Among all the machine learning approaches such as Support Vector Machines, Back Propagation neural network [11] and decision tree, CNN is considered as the smarter classification tool as it can automatically extract the features without picking manually by us. Inspiring by the concept of cats' biological receptive field concept [12] the convolution and pooling layers are present in CNN and those layers are not present in the Back Propagation neural networks, it is the only one difference that is considered between them.

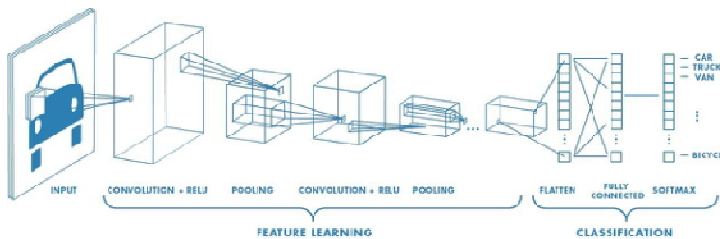


Figure 2 : showing the working of CNN classification

2.6.2 - DPCA

It is the generalized form of Principal Component Analysis, the difference between PCA and 2DPCA is that the 2DPCA is based on the 2-dimensional matrices. There is no need to transform the 2- dimensional matrix into a 1-dimensional vector. So, the 2DPCA has two benefits that are the more precise and time-consuming analysis of the covariance matrix. Let a be the column vector which is Q - dimensional. The 2DPCA projects an image I of $m \times n$ to a , and it obtains a P -dimensional projected vector b . i.e. $b = Ia$.

The representation of Gabor filters is optimal and improves the value of recognition of facial expressions. In the 1980s, Daugman first introduced the use of a 2 - dimensional Gabor Wavelet computer vision. In this article, we use Cheng Jun Liu's paper's Gabor wavelets.

The face recognition which is in the form of 3 – dimensional is like a facial recognition technology that uses the faces of humans which are in the 3 – dimensional structure. Through checking the structure of rigid features on the face, 3D face recognition methods have been shown to achieve significantly higher accuracy rivaling the detection of fingerprints than their

2D counterparts. It prevents 2D face recognition algorithms. Another approach is by making the face an accepted view to use the 3D template to improve image accuracy.

However, as well as the corresponding surface, most 3D scanners achieve a 3D mesh. This makes it possible to combine the efficiency of pure 3D matches with the more conventional 2D face recognition algorithms, resulting in improved performance.

The processing of a 3 – dimensional image that requires a scope of the camera is the key technical drawback of the face recognition methods of 3- dimensional. Instead, with minimal post-processing, multiple images from different angles from a typical camera can be used to construct the 3D model. This is also an explanation of why applications for 3D face recognition have evolved significantly. Recently commercial technologies introduced a profound perception by extending outwards a grid on to the head and incorporating the capture of a video of the high resolution of a 3-dimensional model. A good and highest accuracy is allowed in this with the identification of off-the-shelf components that are of low cost. Though many companies are providing the solutions for business, the research of the face recognition of a 3-dimension is still considered as an active field.

2.7 AdaBoost

Yoav Freund and Robert Schapire are designing Adaptive Boosting. It can be used to improve performance. Certain learning algorithms output is combined into a weighted sum representing the final output. We can adapt the AdaBoost in such instances where the previous classifiers misclassified the subsequent poor learners. AdaBoost is sensitive to noisy data and outliers. In some cases, the problem of overfitting may be less sensitive. The individual learners may be poor, but until the output is far better than the random guesses, it can be shown that the final model converges to a strong learner. A learning algorithm usually has many different parameters and configurations to change until the optimum output is achieved. Adaboost is referred to as the strongest classifier which is considered as the out-of-the-box in which the decision trees are considered as the soft learners. When it is used along with the decision tree learning, after collecting all the information on the sample complexity of the training dataset from every stage of the algorithm of the AdaBoost, that information is given to the tree growing algorithm as it later focuses more on the complex instances to identify.

2.8 Support Vector Machines

It is used in various tasks to recognize the patterns. SVM is an innovative approach to machine learning based on the modern theory of statistical learning. We can achieve the near-optimum by using the SVM among the separation of classes.SVMs are qualified to identify facial expressions using the proposed characteristics. In general, SVM is the maximal hyperplane classification method that relies on results from statistical learning theory to guarantee high generalization performance. Kernel functions are used to model data input

efficiently. SVMs show good classification accuracy even when only a small amount of training data is available, making it particularly suitable for a complex, interactive approach to expression recognition. An ideal separation is achieved when the hyperplane and the training data of any class are the largest. This separating hyperplane works as the decision surface. For a number of classification tasks, SVM has been successfully employed. Below given is the labeled equation of a training dataset.

$$Z = \{(x_k, y_k) \mid x_k \in \mathbb{R}^n, y_k \in \{-1, 1\}\} \quad p \text{ k}=1$$

Here the hyperplane is discovered by the SVM to determine the samples of training dataset with the errors which are very small and hard to detect.

$$e. x - a = 0$$

we will achieve the classification of x_k after computing the distance between the input vector (x_k) and the hyperplane. The SVM is the classifier of the binomial.

2.9 Softmax Classifier

Through mathematics, the softmax function is one that receive input and normalizes, vector of K real numbers into a distribution of probabilities consisting of K probabilities relative to the exponentials of input numbers. This implies that before the application of the softmax classifier function, there may be negative components of vectors or the components of vectors which are greater than one but not add up to 1. The components of vectors vary in the range of 0 to 1 that is it is in the interval (0,1) only after applying the softmax classifier function. The inputs which are having the large components will correspond to the highest probabilities. Over the groups of outputs that are expected, we can use the softmax classifier function in the neural networks to map the output of the network which is not normalized to the distribution of probability.

The default softmax function (unit)

$\sigma: \mathbb{R}^K \rightarrow \mathbb{R}^K$ is the defined equation

$$\sigma(\mathbb{Z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \text{ for } i = 1, \dots, K \text{ and } z = (z_1, \dots, z_K) \in \mathbb{R}^K$$

3. PARAMETERS USED IN THE OPTIMIZERS

3.1 SGD

Keras.optimizers :

SDG learning_rating=0.001,momentum=0.0,nesterov=false

Stochastic gradient descent optimizer.

- learning_rate: float >= 0.
- momentum: float >= 0.

3.2 RMSProp

SDG(learning rating=0.001,rho=0.9).

Arguments

- speed of learning: float >= 0. The speed of reading.
- Rho: >= 0 float.

3.3 Adam

Adam(learning rate=0.001, beta 1=0.9, beta 2=0.9, amsgrad= False)

Arguments

- speed of learning: float >=0. Learning rate.
- Beta 1: 0 <<<< less than 1. Usually close to 1.
- Beta-2:0<<<<<1. Usually close to 1.

Table 3 : Showing the association between optimizers and accuracy of classifiers

Optimizers	Adam	Sgd	Rmsprop
Max.Accuracy	92.09	86.23	91.7
Ultimate.Accuracy	91.21	84.28	89.4

The optimizer plays an important role in the accuracy of classifier. Choosing a right classifier is very important in improving the classification accuracy. Different optimizers are available in the literature and these optimizers are used to test on these images. RMSProp, RMSProp + momentum, Adam optimizer, Standard gradient descent are tested on the images. In this paper adam optimizer and softmax classifier are used to get the better accuracy and the accuracy found out to be 94.82%.The different methods used in the classification are 2D PCA, Adaboost, SVM and the proposed Softmax Classifier. Our method is also compared with other different models. Inorder to get this We have increased the convolutional layers Conv2D with filters 64and 128 for each iterations. Along with it Face recognition is also done.

Table 4 : Showing the Accuracy of different methods

Methods	Accuracy
2D PCA[2]	78.61%
Adaboost[2]	83.12%
SVM[2]	87.15%
PHOG+CLDA[4]	85.34 %
CNN [2]	93.16 %
Softmax Classifier(proposed method)	94.82%

Table 5 : Table showing different emotions and the percentage of emotions detected when input is given with emotion

Labels	Angr y	Disgus t	Fear	Happy	Sad	Surprise	Neut ral
Angry	76.0	7.6	4.9	0.0	4.9	2.4	4.9
Disgus t	5.4	93.9	0.0	0.0	0.0	0.0	0.0
Fear	4.3	0.0	64.9	8.8	0.0	12.9	8.7
Happy	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Sad	12.5	3.9	3.9	0.0	67.9	4.0	7.9
Surpris e	0.0	0.0	0.0	0.0	4.0	95.9	0.0
Neutral	3.3	3.3	0.0	6.4	31.	0.0	84.4

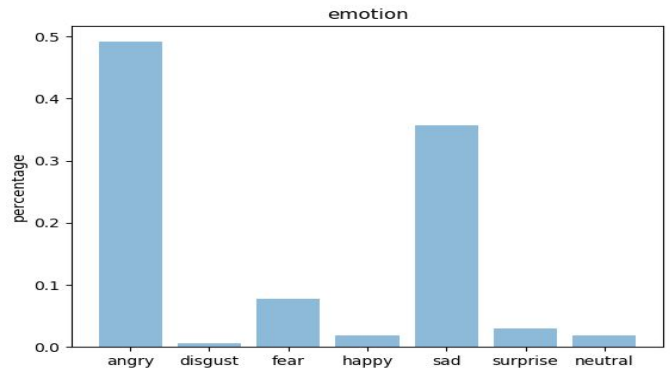


Figure 4 : Graph showing the percentage of other emotions when the input expression is Angry

Table 6 : Table representing the Dataset

Labels	Ang ry	Disgu st	Fea r	Hap py	Neut ral	Sad	Surp rise
Angry	259	0	0	0	0	1	0
Disgust	1	182	0	0	0	0	0
Fear	2	1	219	0	0	0	1
Happy	25	40	173	98	1	19	0
Neutral	1	1	12	0	111	0	0
Sad	1	1	1	1	0	228	0
Surpris e	12	15	141	1	0	11	60

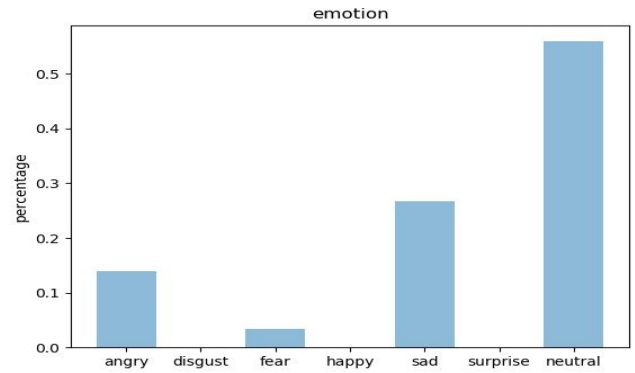


Figure 5 : Graph showing the percentage of other emotions when the input expression is Neutral

The emotion detection is done and also face recognition is also done and the figure shows an example of it.

5.RESULTS

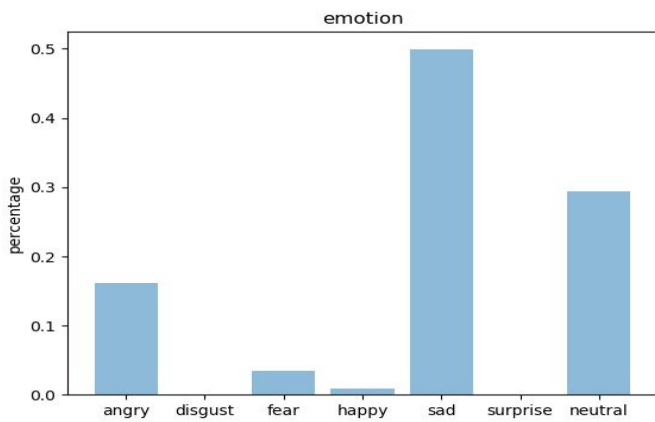


Figure 3 : Graph showing the percentage of other emotions when the input expression is SAD

Face detection

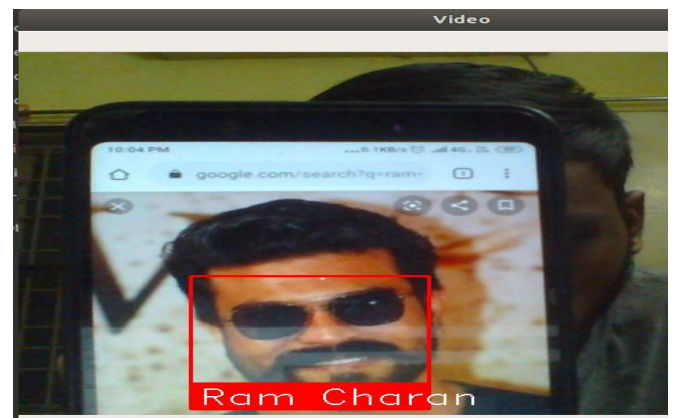


Figure 6 : Showing an example of recognizing a face and displaying the result

6.CONCLUSION

The improvement in the Artificial intelligence, Machine Learning and Deep Learning and their applications in various fields such as robotics, surveillance, latest cameras, safety, Interactions between machines and humans had made to do an

extensive research as they require recognition of the faces and their expressions also their features. The main objective of this paper is to demonstrate a facial expression recognition system to execute and improve the advanced extraction and to determine the classification of facial expression recognition features. In an whole 7 facial expressions are derived from separate data sets for this task. The features of the facial images are extracted using Local Binary Patterns and CNN is used for classification using the datasets. In research it was found that there are many optimizers are available and they are tested, the peak accuracy of the projected technique reaches 94.82 percent, which is an enormous enhancement compared to other techniques.

REFERENCES

1. E.Sariyandi, H.Gunes. and A. Cavallaro. **“Automatic analysis of facial affect: A survey of registration, representation and recognition,”** *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, num.6, pp. 1113-1133, 2015.
2. Danyang Qu, Zheng Huang, Zhenyuan Gao, Yiwen Zhao, Xin'gang Zhao, Guoli Song **“An Automatic System for Smile Recognition Based on CNN and Face Detection”** *Proceedings of the 2018 IEEE International Conference on Robotics and Biomimetics* December 12-15, 2018, Kuala Lumpur, Malaysia
3. Huiyuan Yang, Lijun Yin, **“CNN based 3D Facial Expression Recognition Using Masking and Landmark Reatures,”** *International Conference on Affective Computing & Intelligent Interaction*, San Antonio, 2017, pp.556-560.
4. L.H. Guo, Y. Bai, L.W. Jin. **“Smile Recognition Based on PHOG Feature Extraction and Clustering Feature Selection,”** *Pattern Recognition & Artificial Intelligence*, vol. 25, num. 1, pp.23-28, 2012.
5. Xiru Wu'et al. **“Fast Visual Identification and Location Algorithm for Industrial Sorting Robots”** *Based on Deep Learning [J]. ROBOT*, 2016, 38(6):711-719.
6. **“ The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression”** Patrick Lucey^{1,2}, Jeffrey F. Cohn^{1,2}, Takeo Kanade¹, Jason Saragih¹, Zara Ambadar² *Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 152131 Department of Psychology, University of Pittsburgh, Pittsburgh, PA, 152602*
7. M. Bindemann, A. M. Burton. **“The role of color in Human Face Detection,”** *Cognitive Science*, vol. 33, num. 6, pp. 1144-1156, 2009. <https://doi.org/10.1111/j.1551-6709.2009.01035.x>
8. P. Viola, M. Jones. **“Rapid object detection using a Boosted Cascade of simple features,”** *IEEE Computer Society Conference on Computer Vision & Pattern Recognition, Madison, 2003*, pp. 511-518.
9. T. Roska, L. O. Chua. **“The CNN universal machine: an analogic array computer,”** *IEEE Transactions on Circuits & Systems II Analog & Digital Signal Processing*, vol. 40, num. 3, pp. 163-173, 2015.
10. Y. Lecun, K. Kavukcuoglu, C. Farabet. **“Convolutional networks and applications in vision,”** *IEEE International Symposium on Circuits and Systems*, vol. 14, num. 5, pp. 253-256, 2011.
11. Zhiyu Zhou' Ming Yang' et al. **“Vegetation Detection Approach Based on Gaussian Kernel Support Vector Machine”** *in Unstructured Road Environment [J]. ROBOT*, 2015, 37(6):702-707.
12. W. Shi, J. Caballero, F. Huszar, et al. **“Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network,”** *IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016*, pp. 1874-1883.
13. Vamsidhar Enireddy and Kiran K Reddi **“A Data Mining Approach for Compressed Medical Image Retrieval”** *in IJCA International Journal of Computer Applications* Volume 52– No.5, August 2012.
14. Enireddy Vamsidhar and Kiran K Reddi **“Improved Cuckoo Search with Particle swarm Optimization for Classification of Compressed Medical Images”** *Sadhana* Vol. 40, Part 8, December 2015, pp. 2271–2285. *Indian Academy of Sciences*.
15. Vamsidhar Enireddy, Sai Chandana and HariKiran j **“A Novel Approach for Feature selection and Classifier Optimization Compressed Medical Retrieval Using Hybrid cuckoo Search”** *Indonesian Journal of Electric Engineering and Information* Vol.6, No.4, December 2018, pp.410-417.
16. Prakash, K.B. & Rajaraman, A. 2016, **“Mining of Bilingual Indian Web Documents”**, *Procedia Computer Science*, Volume 89, 2016, Pages 514-520. <https://doi.org/10.1016/j.procs.2016.06.103>
17. Prakash, K.B. & Rangaswamy, M.A.D. 2016, **“Content extraction of biological datasets using soft computing techniques”**, *Journal of Medical Imaging and Health Informatics*, Volume 6, Issue 4, August 2016, Pages 932-936.
18. Prakash, K.B. 2015, **“Mining issues in traditional indian web documents”**, *Indian Journal of Science and Technology*, vol. 8, no. 32, pp. 1-11.
19. Vamsidhar Enireddy, Karthikeyan C, Manikandan J **“Image Fusion using Mallat Algorithm of Wavelet Transform”** *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-8 Issue-11, September 2019. <https://doi.org/10.35940/ijitee.K2040.0981119>
20. Sai Narendra L, Samuel Kiran, Naga Brahmani K. Vamsidhar E **“ Performance Analysis on Human Activity Detection using KNN and Random Forest”** *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-8 Issue-7 May, 2019