iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44396

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Music Recommendation System Using Facial Expression Recognition Using Machine Learning

B. Nareen Sai[1], D. Sai. Vamshi[2], Piyush Pogakwar[3], V. Seetharama Rao[2], Y. Srinivasulu[3]
[1, 2, 3]*Students,* [4]*Asst. Professor,* [5]*Assc. Professor Department of Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology Technology, Yamnampet, Hyderabad, Telangana, India*

Abstract: ***"MUSIC RECOMMENDATION SYSTEM USING FACIAL EXPRESSION RECOGNITION USING MACHINE LEARNING" is the title of this project. The study of human emotional responses to visual stimuli such as photos and films, known as visual sentiment analysis, has proven a fascinating and difficult problem. It attempts to comprehend the high-level information of visual data. The development of strong algorithms from computer vision is responsible for the present models' success. The majority of existing models attempt to overcome the problem by recommending either robust features or more sophisticated models. Visual elements from the entire image or video are the key proposed inputs in particular. Local areas have received less attention, which we believe is important to the emotional reaction of humans to the entire image. Image recognition is used to locate people in photos, analyse their emotions, and play emotion-related tunes based on their feelings. This repo accomplishes this goal by utilising the Google platform's Vision services. Given an image, it would look for faces, identify them, draw a rectangle around them, and describe the emotion it discovered.***

Keywords: ***Emotion Recognition; music recommendation; Facial Extraction; Emotion Extraction Module; Audio Feature Extraction Module***

I. INTRODUCTION

When they have hundreds of tracks, music listeners find it difficult to manually create and organise playlists. It's also tough to keep track of all the music: some are added but never played, wasting a lot of device capacity and forcing the user to manually locate and delete songs. Users must actively choose music depending on their interests and mood each time. When users' play-styles change, it's also tough to reorganise and play music. As a result, we employed the Machine Learning idea, which entails face scanning and feature monitoring, to assess the user's mood and create a tailored playlist based on it.



Fig -1: Various basic emotions of humans

II. LITERATURE SURVEY

S Metilda Florence and M Uma (2020) proposed a paper "*Emotional Detection and Music Recommendation System based on User Facial Expression*" When a user's facial expressions are detected by the system, it extracts facial landmarks that may be used to classify the user's emotional state. A list of songs that best express the user's current mood will be shown to him or her after the emotion has been identified. It might help a user make a decision on what music to listen to in order to lessen their stress levels. Searching for tunes would be a thing of the past for the user. There were three components in the suggested architecture: Emotion-Audio extraction module, Audio extraction module, and Emotion-Emotion extraction module. As a result of a smaller picture dataset being employed, this suggested system wasn't able to accurately record all of the different emotions that people felt. In order for the classifier to produce reliable results, the picture input should be taken in a well-lit environment. For the classifier to properly estimate the user's emotional state, the image quality must be at least 320p. In the real world, the generalizability of handcrafted features is frequently lacking.

©IJRASET: All Rights are Reserved

H. Immanuel James, J. James Anto Arnold, J. Maria Masilla Ruban, M. Tamilarasan (2019) proposed "*Emotion Based Music Recommendation*" It seeks to scan and analyse facial expressions before constructing a playlist based on them. By producing a suitable playlist based on an individual's emotional traits, the tiresome work of manually segregating or dividing music into distinct lists is minimised. The suggested method focuses on sensing human emotions in order to create music players that are emotion-based. Face detection is done with a linear classifier. Using regression trees trained with a gradient boosting approach, a facial landmark map of a given face picture is constructed based on the pixel intensity values indexed of each point. Emotions are classified using a multiclass SVM Classifier. Emotions are divided into four categories: happy, angry, sad, and surprised. The suggested system's shortcomings include the fact that due to the limited availability of images in the image dataset employed, it is currently unable to accurately record all emotions. There are no diverse feelings to be found. In the wild, handcrafted characteristics are frequently insufficiently generalizable.

Ali Mollahosseini, Behzad Hasani and Mohammad H. Mahoor (2017) proposed *"AffectNet"* : More than 1,000,000 facial images were obtained from the Internet by querying three major search engines using 1250 emotion-related keywords in six different languages, according to "A Database for Facial Expression, Valence, and Arousal Computing in the Wild," where more than 1,000,000 facial images were obtained from the Internet by querying three major search engines using 1250 emotion-related keywords in six different languages. The existence of seven unique facial expressions, as well as the strength of valence and arousal, were manually annotated in around half of the recovered photos. In the continuous domain of the dimensional model, two baselines are proposed to categorise pictures in the categorical model and predict the value of valence and arousal. There were certain restrictions, such as the fact that VGG16 only improves on AlexNet by replacing huge kernel-sized filters with multiple 3X3 kernel-sized filters one by one. Multiple stacked smaller size kernels are better than one bigger size kernel in a given receptive area. The samples in the AffectNet database aren't particularly powerful. That is, samples with arousal of 1 or -1 and valence of 1 or -1.

III. PROPOSED SYSTEM

Convolutional Neural Networks (CNN) are used to learn the most relevant feature abstractions directly from data collected in an uncontrolled environment, as well as to deal with the limits of handmade features. CNNs have been proven to be a successful method for visual object detection, human position estimation, and face verification, among other applications. DNNs can extract highly discriminative features from data samples thanks to the availability of computer power and large databases. CNNs have been shown to be extremely successful in picture recognition and classification. CNNs are exceptionally good at lowering the amount of parameters without sacrificing model quality.

Using the VGG16 CNN model, the suggested system can recognise the user's facial expressions and make decisions based on such expressions. After the emotion has been categorised, a music that corresponds to the user's feelings will be played.

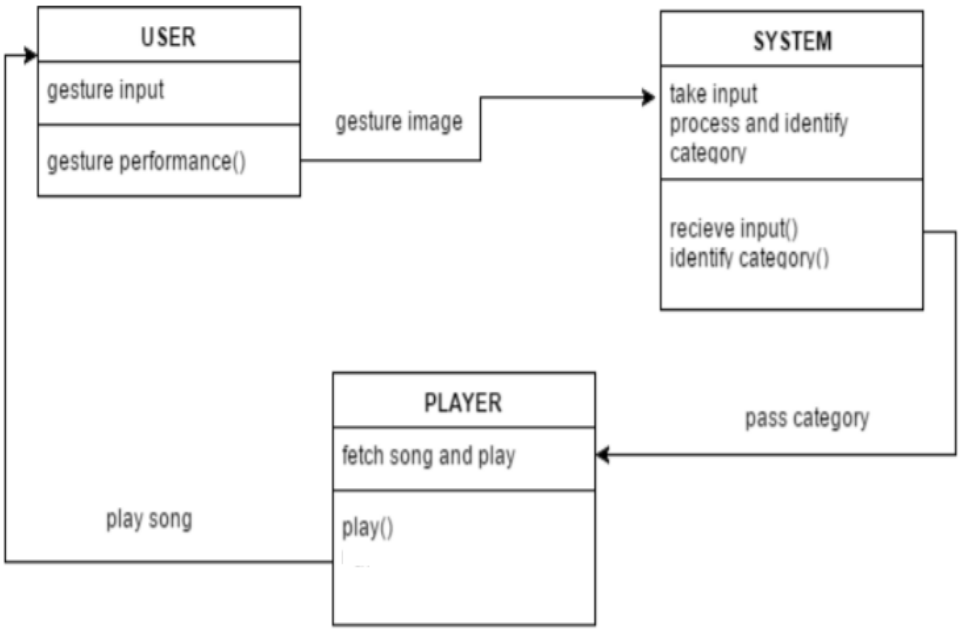


Fig -2: Proposed System

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

IV. SYSTEM ARCHITECTURE

We explain our suggested system for analysing students' facial expressions using a Convolutional Neural Network (CNN) architecture. First, the system finds faces in the input image, which are then cropped and normalised to a size of 4848 pixels. These photos of people's faces are then transmitted into CNN. Finally, the face expression recognition results are shown (anger, happiness, sadness, disgust, surprise or neutral). Figure 5.1 depicts the framework of our suggested strategy. When compared to other image classification techniques, a Convolutional Neural Network (CNN) is a deep artificial neural network that can recognise visual patterns from input photos with minimum preparation. This means that the network picks up on the filters that were previously hand-crafted in traditional methods [19]. Within a CNN layer, the most significant unit is the neuron. They're connected in such a way that the output of one layer's neurons becomes the input of the following layer's neurons. The partial derivatives of the cost function are computed using the backpropagation technique. Convolution is the process of creating a feature map from an input picture by applying a filter or kernel. In reality, the CNN model contains three types of layers, as shown in Figure 3.

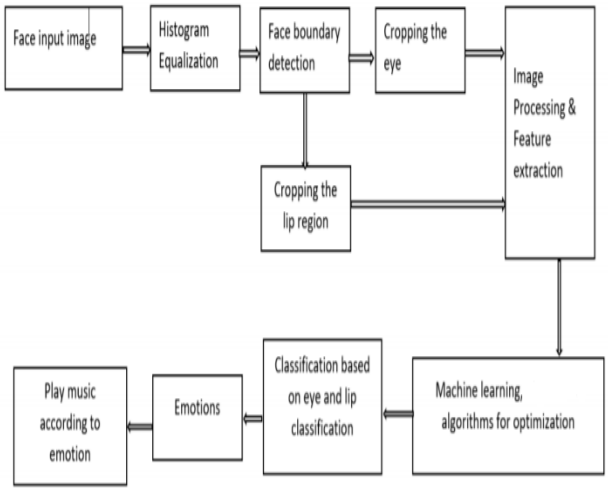


Fig -3: System Architecture of Music Recommendation System using facial expression recognition

A. *Modules*

1) *User*: With this module, a user may connect a camera to an application and track live video of a user with various emotional expressions, as well as take images of each frame and transmit data to a CNN model that will forecast and provide live emotions. Based on the studied data values provided from the CNN model and received by the user, the live camera was shown with the type of sentiment (sad, happy..etc)
2) *Dataset Collection*: Collect data from the 2013 dataset, which includes pixlel values as features and emotions as labels.
3) *Pre processing:* This stage divides the dataset into features and labels, which are then stored as x and y values.
4) *Initializing CNN*: The CNN model is initialised at this step, and features and labels are provided to the fit statement, and the algorithm is trained. The model is stored as a.h5 file.
5) *CNN Model Training*: Fer 2013 data set with facial emotions is used as input, and a CNN algorithm model is developed and saved to disc for prediction of live emotions. ( )
6) *Prediction*: At this point, the camera will open, and faces will be detected and forecasted using a trained model, with an emotion song being played based on the results.

V. IMPLEMENTATION

We gathered data for our CNN architecture using the FER2013 [12] database. It was constructed using the Google image search API and presented at the 2013 ICML Challenges. Faces in the database were automatically adjusted to 4848 pixels. There are 35887 pictures in the FER2013 collection with 7 expression labels (28709 training images, 3589 validation images, and 3589 test images). A fixed amount of photos are used to symbolise each emotion. CNN is a cable news network (CNN) Putting everything together: We utilised the OpenCV package to record live web camera shots and apply the Haar Cascades approach to recognise students' faces. The Adaboost learning method is used in Haar Cascades.

©IJRASET: All Rights are Reserved

The Adaboost learning method picked a small number of relevant features from a vast set in order to provide an effective classifier output. To do picture augmentation, we utilised Keras' Image Data Generator class. We were able to rotate, shift, shear, zoom, and flip the training photos using this class. Rotation range=10, width shift range=0.1, zoom range=0.1, height shift range=0.1, and horizontal flip=True are the parameters utilised. Then we employed four convolutional layers, four pooling layers, and two fully linked layers in our CNN model. The ReLU function was then used to add nonlinearity to our CNN model, as well as batch normalisation to normalise the activation of the previous layer at each batch and L2 regularisation to apply penalties to the model's different parameters. 22

We used the FER 2013 database to train our Convolutional Neural Network model, which covers seven emotions (happiness, anger, sadness, disgust, neutral, fear and surprise) Before being input into the CNN model, the identified face photos were scaled to 4848 pixels and transformed to grayscale images.

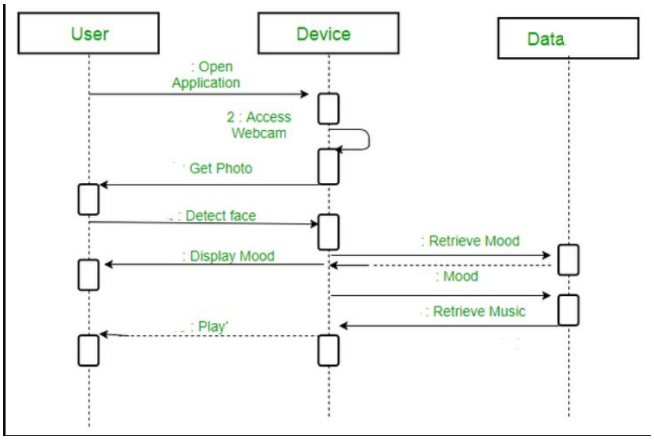


Fig-4: Implementation of Music Recommendation System using facial expression recognition.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

The emotion recognition of the photographs fed into the suggested model is the most important aspect of this study. The main goal is to use the emotion detecting feature. The suggested approach aims to improve an individual's entertainment by integrating emotion recognition technologies with a music player.

The proposal can detect four different emotions: normal, joyful, and sad. When the suggested model detects an emotion, the music player will play the appropriate song(s). In terms of usefulness and accuracy, the suggested model has undergone both system and emotion accuracy testing, with positive results. The suggested model was able to recognise 34 of the 40 photos fed into it, giving it an 85 percent recognition rate. Furthermore, the suggested model is a computer programme that may run on a variety of operating systems and machines.

As a result of our Emotion Based Music Player, users may choose music in a more participatory and straightforward manner. It can assist music fans in automatically searching for and playing songs based on their feelings.

B. Recommendation

Every system, including the Emotion Based Music Player, is subject to modifications and improvements.

The first step is to improve emotion detection accuracy. It is possible to do this by expanding the amount of face characteristics utilised in emotion recognition. At the moment, the model simply extracts the mouth and eyes. Other face characteristics, such as brows and cheeks, might be introduced in future development. Furthermore, noise reduction software may be included in the model to automatically eliminate noise from still or recorded images. Apart from the foregoing, the suggested model might be enhanced by include auto-adjustment of picture resolution, brightness, and contrast. The quality of the photographs loaded has a significant impact on the accuracy of emotion recognition in the current application. As a result of the auto adjustment, the user may load any image quality, and the future model will be able to modify the quality to meet its needs.

Furthermore, a real-time emotion recognition technology may be used to improve the interaction between the user and the programme. Once the app is activated, the future model will identify and extract face features, allowing the emotion to be identified in real time.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VII. ACKNOWLEDGEMENT

First and foremost, the writer wishes to express theIr heartfelt gratitude and admiration to the project's principal supervisor and internal guide, Asst. Prof. Mr. V Seetharama Rao, who kept a close eye on the project's development throughout the semesters. The project's success has been attributed to the constructive remarks and recommendations supplied.

This thanks extends to anyone who took the time to fill out the survey. The author is grateful for the input and collaboration provided by the participants, which has aided in the development and enhancement of the system prototype.

Last but not least, the author would like to thank her family and friends for their unwavering support and constructive criticism throughout the process.

REFERENCES

[1] Savva, V. Stylianou, K. Kyriacou, and F. Domenach, "Recognizing student facial expressions: A web application," in Tenerife, 2018 IEEE Global Engineering Education Conference (EDUCON), p. 1459-1462.

[2] D. Bahdanau, K. Cho, and Y. Bengio. 2014. By learning to align and translate together, neural machines can translate. Presented at ICLR 2015.

[3] Borth, D., Chen, T., Ji, R., and Chang, S.-F. Sentibank: large-scale ontology and classifiers for recognising sentiment and emotions in visual information.

[4] Umesh Kumar, Jagdish Lal Raheja Human Facial Expression Detection From Detected In Captured Image Using Back Propagation Neural Network, International Journal of Computer Science & Information Technology (IJCSIT), Vol.2 (2010). (1).

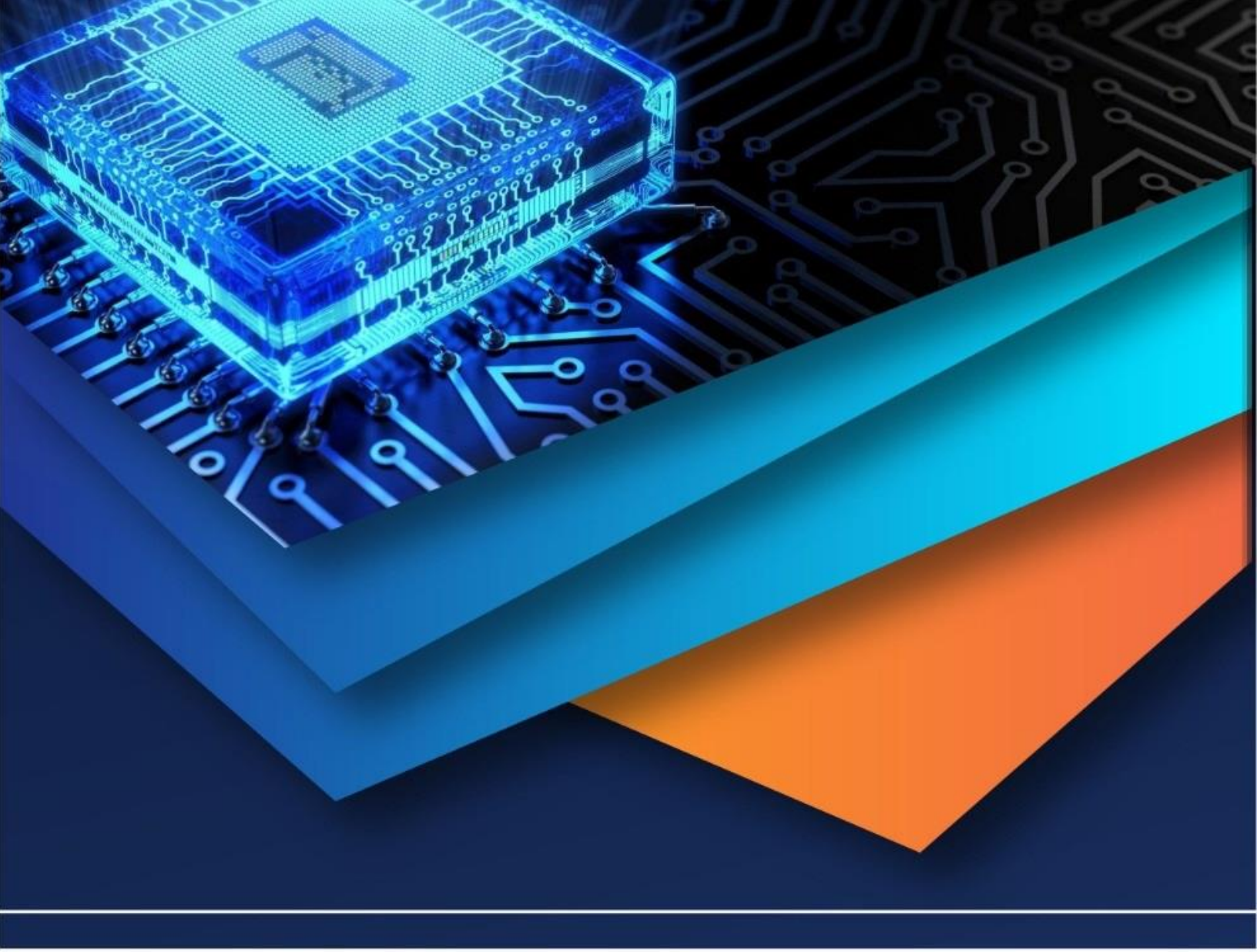
[5] Zhengyou Zhang, Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiments With a Multi-Layer Perceptron, Journal of Pattern Recognition and Artificial Intelligence, vol. 13(6), pp. 893- 911, 1998.

[6] Eva Cerezo1, Isabelle Hupont, Critina Manresa, Javier Varona, Sandra Baldassarri, Francisco J. Perales, and Francisco J. Seron. Eva Cerezo1, Isabelle Hupont, Critina Manresa, Javier Varona, Sandra Baldassarri, Francisco J. Perales, and Francisco J. Seron. Real-Time Facial Expression Recognition for Natural Interaction, in J. Mart et al. (Eds.), IbPRIA 2007, Part II, LNCS 4478, pp. 40–47. Heidelberg: Springer-Verlag Berlin.

[7] C. Shan, S. Gong, and P. W. McOwan, "A thorough investigation of facial emotion identification based on local binary patterns," Image Vis. Comput., vol. 27, no. 6, pp. 803–816, May 2009.

[8] Antoinette L. Bouhuys, Gerda M. Bloem, Ton G.G. Groothuis. Antoinette L. Bouhuys, Gerda M. Bloem, Ton G.G. Groothuis. Induction of depressed and elated moods by music influences the perception of facial emotional expressions in healthy subjects: Journal of Affective Disorders, vol. 33, no. 2, pp. 215-226.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)