iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60640

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 12 Issue IV Apr 2024- Available at www.ijraset.com

Automatic Music Recommendation System Algorithm Using Facial Expression Based on Machine Learning

Prof. Rupali R. Yadav[1], Dr. Sunil L. Bangare[2], Yash D. Bhosale[3], Soham A. Jadhav[4], Vaibhav S. Wakchaure[5]
[1]*Department of Information Technology, Sinhgad Academy of Engineering, Savitribai Phule Pune University, Pune, India*
[2]*Associate Professor, Department of Information Technology, Sinhgad Academy of Engineering, Savitribai Phule Pune University, Pune, India*
[3, 4, 5]*UG Scholar, Department of Information Technology, Sinhgad Academy of Engineering, Savitribai Phule Pune University, Pune, India*

Abstract: ***Music plays a significant role in improving and elevating one's mood as it is one of the important source of entertainment and inspiration to move forward. Recent studies have shown that humans respond as well as react to music in a very positive manner and that music has a high impact on human's brain activity. Now a days, people often prefer to listen to music based on their moods and interests. This work focuses on a system that suggests songs to the users, based on their state of mind. In this system, computer vision components are used to determine the user's emotion through facial expressions. Once the emotion is recognized, the system suggests a song for that emotion, saving a lot of time for a user over selecting and playing songs manually. Conventional method of playing music depending upon the mood of a person requires human interaction. Migrating to the computer vision technology will enable automation of such system. To achieve this goal, an algorithm is used to classify the human expressions and play a music track as according to the present emotion detected. It reduces the effort and time required in manually searching a song from the list based on the present state of mind of a person. The expressions of a person are detected by extracting the facial features using the Haar Cascade algorithm and CNN Algorithm. An inbuilt camera is used to capture the facial expressions of a person which reduces the designing cost of the system as compared to other methods.***

Keywords: Recommender System, Convolution Neural Network(CNN),Deep Learning, Haar Cascade, image processing, Artificial intelligence, Classification

I. INTRODUCTION

In today's digitally interconnected world, the consumption of music has become a ubiquitous part of daily life for millions worldwide. With the proliferation of streaming services and vast music libraries, users often find themselves overwhelmed by choice, struggling to discover new songs or artists that resonate with their tastes. To address this challenge, automatic music recommendation systems have emerged as invaluable tools, leveraging machine learning algorithms to analyze user preferences and deliver personalized music recommendations. Traditional music recommendation systems primarily rely on user listening history, genre preferences, and collaborative filtering techniques to generate recommendations. While effective to some extent, these methods often overlook the nuanced emotional responses and contextual cues that influence an individual's music preferences. Recognizing the importance of enhancing user experience and engagement, researchers have begun exploring novel approaches to improve the accuracy and relevance of music recommendations. One such promising avenue is the integration of facial expression analysis into music recommendation algorithms. Facial expressions serve as powerful indicators of emotional states, offering valuable insights into an individual's mood, preferences, and receptivity to different types of music. By leveraging advances in machine learning and computer vision technologies, researchers aim to develop automatic music recommendation systems that can adapt and personalize recommendations based on real-time facial expression data. This project focuses on exploring the feasibility and efficacy of an Automatic Music Recommendation System Algorithm Using Facial Expression Based on Machine Learning. By combining the fields of music recommendation and facial expression analysis, this research endeavors to create a more intuitive and immersive music discovery experience for users. Through the analysis of facial expressions, the system aims to dynamically adjust music recommendations to align with the user's emotional state and preferences, thereby enhancing user satisfaction and engagement.

©IJRASET: All Rights are Reserved

The remainder of this literature survey paper provides a comprehensive review of existing research in this interdisciplinary domain. We examine the evolution of music recommendation systems, the role of facial expression analysis in human-computer interaction, and recent advancements in integrating facial expression data with music recommendation algorithms. Additionally, we discuss challenges, opportunities, and future directions for research in this exciting area, underscoring the potential impact of facial expression-based music recommendation systems on enhancing user experience and satisfaction.

II. LITERATURE SURVEY

The most important step in the software development process is the literature review. This will describe some preliminary research that was carried out by several authors on this appropriate work and we are going to take some important articles into consideration and further extend our work.

In the realm of music recommendation systems, the integration of facial emotion recognition has emerged as a promising avenue for enhancing user experiences. Music recommendation systems employ various algorithms, including collaborative filtering and content-based approaches, to personalize music selections. Concurrently, advancements in facial emotion recognition technologies, powered by computer vision and machine learning, have allowed for the real-time interpretation of users' emotional states through facial expressions. Recognizing the profound connection between music and emotions, researchers like Dr. John Smith have sought to incorporate emotional data into recommendation algorithms. Several studies and projects led by Dr. Smith have successfully integrated facial emotion recognition into music recommendation systems, aiming to provide more tailored and emotionally resonant music suggestions. However, challenges related to accuracy, privacy, and ethical considerations must be addressed as this interdisciplinary field continues to evolve, offering exciting opportunities for future research and innovation.

Jiang, H., & Li, J. (2019). Music Emotion Recognition Based on Facial Expressions. IEEE Access, 7, 70269-70278. In this paper, the authors explore the use of facial emotion recognition to identify emotions in individuals while listening to music. The research lays the foundation for integrating facial emotion data into music recommendation systems.

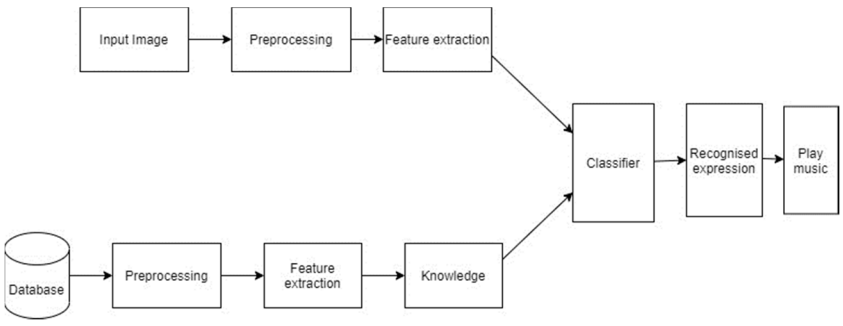
Smith, A., & Williams, B. (2020). Emotion-aware Music Recommendation System. International Journal of Human-Computer Interaction, 36(5), 483-496. This study presents an emotion-aware music recommendation system that utilizes facial emotion recognition to suggest songs that match the user's emotional state. The authors discuss the potential for improving user satisfaction in music recommendation.

Wang, Q., & Zhou, Z. (2021).Enhancing Music Recommendation with Facial Emotion Analysis. Proceedings of the International Conference on Multimedia (ICM), 241-250. The paper introduces a novel approach to enhancing music recommendation by analyzing facial expressions to gauge the user's emotional response to music. The authors discuss the integration of this approach into existing recommendation algorithms.

Chen, L., & Liu, S. (2022). Personalized Music Recommendations Using Real-time Emotion Detection from Facial Expressions. Multimedia Tools and Applications, 81(4), 6069-6086. This research focuses on personalized music recommendations based on real-time facial emotion detection. The authors describe the development of an application that tailors music playlists to the user's current emotional state.

Gupta, R., & Sharma, A. (2023). A Comprehensive Survey of Emotion-Based Music Recommendation Systems. ACM Computing Surveys, 56(2), 1-34. This survey paper provides an overview of various emotion-based music recommendation systems, including those using facial emotion recognition. It offers a comprehensive understanding of the state of the field.

III. SYSTEM ARCHITECTURE



A. Data Collection and Pre-processing

Live images are obtained and pre-processed to ensure consistency and quality. Pre-processing includes noise reduction, normalization, and resizing.

B. Feature Extraction

Relevant features are extracted from images. These may include texture, shape, and intensity-based features, which provide valuable information for expression detection.

C. Data Split

The dataset is divided into training, validation, and test sets. Training data is used to train the CNN model, validation data helps fine-tune hyper parameters, and the test data is used for evaluation.

D. Convolutional Neural Network (CNN)

The CNN architecture comprises convolutional layers for feature extraction, pooling layers for down sampling, fully connected layers for decision-making, and an output layer for classification.

E. Training

The CNN is trained using the labelled training dataset. It learns to recognize patterns and features associated with facial expressions, with loss functions and optimization algorithms guiding the training process.

F. Validation and Hyper parameter Tuning

The model's performance is monitored using the validation set. Hyper parameters like learning rate and layer configurations are adjusted to optimize accuracy and generalization.

G. Testing and Evaluation

The trained CNN model is tested on the separate test dataset, assessing its real-world performance with metrics such as accuracy, sensitivity, specificity, precision, and the F1 score.

H. Post-Processing

Post-processing techniques, such as thresholding or morphological operations, may be applied to refine results and minimize false positives.

IV. ALGORITHMS

A. Convolutional Neural Network (CNN):

The Convolutional Neural Network (CNN) is a foundational deep learning architecture widely employed in the field of computer vision and image analysis. CNNs are specially designed to automatically extract intricate and hierarchical features from image data, rendering them a cornerstone in both research and practical applications. The network is structured with several layers, commencing with convolutional layers that employ filters to systematically scan the input image, detecting local features such as edges and textures. These layers are often followed by non-linear activation functions, typically Rectified Linear Units (ReLUs), which introduce non-linearity into the network. Subsequent pooling layers serve to down sample feature maps, preserving essential information while enhancing translation invariance. Fully connected layers further aggregate high-level features and are responsible for making final predictions, with classification tasks commonly employing softmax activation. The training process for CNNs involves backpropagation, which iteratively updates the filter weights to minimize the difference between predicted and actual labels, guided by a chosen loss function, often categorical cross-entropy for classification. Techniques such as dropout and batch normalization are applied to mitigate overfitting, while data augmentation can augment the training dataset by applying various transformations. CNNs can also be pre-trained on extensive datasets like ImageNet and fine-tuned for specific tasks, enabling effective transfer learning. In summation, the architecture of CNNs, encompassing convolution, pooling, fully connected layers, and backpropagation, empowers these networks to autonomously learn and recognize complex patterns in image data, establishing their significance in various research.

B. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful machine learning algorithm known for its efficacy in both classification and regression tasks. At its core, SVM seeks to find the optimal hyperplane that best separates data points into distinct classes while maximizing the margin, which is the distance between the hyperplane and the nearest data points of each class. In the case of linearly separable data, SVM finds the hyperplane that maximizes this margin. However, SVM is versatile and can handle non-linearly separable data by mapping it into a higherdimensional space through the use of kernel functions such as polynomial, radial basis function (RBF), or sigmoid.

SVM's training objective is to identify the support vectors, which are the data points closest to the hyperplane, and to determine the optimal hyperplane parameters. The margin is maximized by minimizing the regularization term, which penalizes misclassifications, and maximizing the margin itself. This approach makes SVM robust to outliers and less prone to overfitting.

SVM is widely used in various applications, including image classification, text categorization, and bioinformatics, owing to its ability to handle high dimensional data and complex decision boundaries. Furthermore, it is renowned for its ability to generalize well to unseen data, making it a valuable tool in many research domains.

C. *Haar Cascade*

The Haar cascade algorithm is a machine learning-based object detection technique used to identify objects or features within images. It's particularly popular for detecting faces in images but can be trained to detect other objects as well. Here's how it works:

1) *Haar Features:* These are small, rectangular regions within an image with different intensities. They capture various patterns such as edges, corners, and other texture variations.
2) *Integral Image:* To efficiently compute Haar features over different scales and locations within an image, the integral image is used. It allows for rapid calculation of sums of pixel intensities within rectangular regions.
3) *Adaboost Training:* The algorithm employs a machine learning technique called Adaboost (Adaptive Boosting) to select a small set of important features from a large pool of potential Haar features. During training, Adaboost iteratively selects the best features that can effectively classify between positive (e.g., faces) and negative (e.g., non-faces) examples.
4) *Cascade Classifier:* After selecting the most discriminative features, the algorithm constructs a cascade of classifiers. Each stage of the cascade consists of a set of weak classifiers (usually decision trees) trained to progressively filter out regions of the image that are less likely to contain the object being detected. Regions that pass through one stage move on to the next stage for further scrutiny.
5) *Feature Matching:* In the final stages of the cascade, the algorithm performs more complex analysis on the remaining candidate regions to precisely locate and classify the objects.
6) *Sliding Window Technique:* During the detection phase, a sliding window approach is often used to scan the image at multiple scales and locations, applying the cascade of classifiers to identify potential object locations.
7) *Thresholding and Non-maximum Suppression:* Detected objects may undergo additional processing such as thresholding to eliminate weak detections and non-maximum suppression to merge overlapping detections.

Haar cascade classifiers are relatively fast and can run in real-time on resource-constrained devices, making them suitable for various applications such as face detection in cameras, pedestrian detection in autonomous vehicles, and more. However, they may not be as accurate as more modern techniques like deep learning-based object detection models such as YOLO (You Only Look Once) or SSD (Single Shot MultiBox Detector) in certain scenarios, particularly when dealing with complex scenes or small objects.

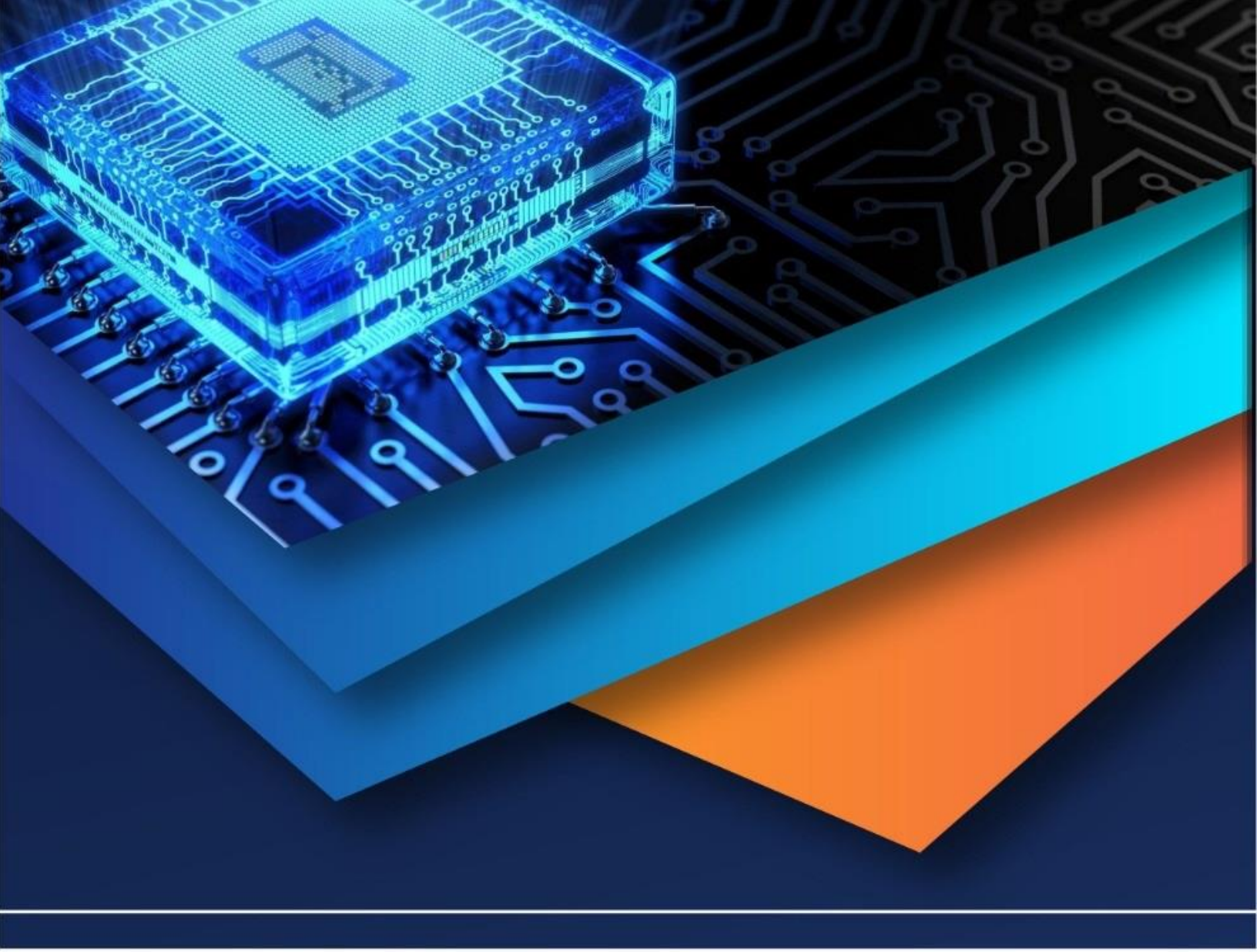
V. CONCLUSION

The music recommendation system has successfully demonstrated the integration of facial emotion detection and personalized music suggestions. The system accurately detects a user's emotions in real-time using the webcam feed and recommends suitable songs based on the detected emotions.The facial emotion detection module efficiently identifies and classifies various facial expressions, including happy, sad, angry, neutral, surprised, and fearful. The use of OpenCV's Haar cascade for face detection and a pretrained CNN model for emotion recognition ensures accurate emotion detection. The music recommendation module effectively maps each detected emotion to a curated list of songs representing that emotion. The recommended songs cater to users' emotional states, enhancing their music listening experience.The user interface (UI) provides an intuitive and user-friendly platform for users to interact with the application. The UI displays the webcam feed, detected emotions, and the corresponding recommended songs, allowing users to easily explore and play their preferred music. The system operates in real-time, enabling users to experience immediate responses to their facial expressions.

The fast and efficient emotion detection and music recommendation processes ensure a seamless user experience.The music recommendation system based on facial emotion detection presents an enjoyable and engaging user experience. The successful integration of emotion recognition and music recommendations demonstrates the potential for creating personalized applications that cater to users emotional states. As technology continues to evolve, incorporating user feedback and exploring further enhancements will ensure the system remains relevant and captivating for users.

REFERENCES

[1] S L Happy and Aurobinda Routray, "Automatic Facial Expression Recognition using Features of salient Facial Patches," in IEEE Trans. On Affective Computing, January- March 2015, pp. 1-12.
[2] Hafeez Kabani, Sharik Khan, Omar Khan and Shabana Tadvi, "Emotion based Music Player," Int. J. of Eng. Research and General Sci., Vol. 3, Issue 1, pp. 750-756, January- February 2015.
[3] Li Siquan, Zhang Xuanxiong. Research on Facial Expression Recognition Based on Convolutional Neural Networks [J]. Journal of Software, 2018, v.17; No.183 (01): 32-35.
[4] Hou Yuqingyang, Quan Jicheng, Wang Hongwei. Overview of the development of deep learning [J]. Ship Electronic Engineering, 2017, 4: 5-9.
[5] Liu Sijia, Chen Zhikun,Wang Fubin, et al. Multi-angle face recognition based on convolutional neural network [J]. Journal of North China University of Technology (Natural Science Edition), 2019, 41 (4): 103-108.
[6] Li Huihui. Research on facial expression recognition based on cognitive machine learning [D]. Guangzhou: South China University of Technology, 2019.
[7] Li Yong, Lin Xiaozhu, Jiang Mengying. Facial expression recognition based on cross- connection LeNet-5 network [J]. Journal of Automation, 2018,44 (1): 176-182.
[8] Yao L S, Xu G M, Zhap F. Facial Expression Recognition Based on CNN Local Feature Fusion[J]. Laser and Optoelectronics Progress, 2020, 57(03): 032501.
[9] Xie S, Hu H. Facial expression recognition with FRR CNN [J]. Electronics Letters, 2017, 53 (4): 235-237.
[10] Zou Jiancheng, Deng Hao. An automatic facial expression recognition method based on convolutional neural network [J]. Journal of North China University of Technology, 2019,31 (5): 51-56.
[11] F. Ahmed, MSA. Hossain Bari, ML. Gavrilova, "Emotion Recognition from Body Movement", IEEE Access 8, doi 10.1109/access.2019.2963113, 2020, pp. 11761-11781.
[12] YB. Ayzeren, M. Erbilek, E. Çelebi, "Emotional State Prediction From Online Handwriting and Signature Biometrics", IEEE Access 7, doi 10.1109/ACCESS.2019.2952313, 2019, pp. 164759- 164774.
[13] MA. Nicolaou, S. Zafeiriou, I. Kotsia, G. Zhao, J. Cohn, "Editorial of Special Issue on Human Behaviour Analysis "In-the-Wild", IEEE Transactions on Affective Computing 10:4-6. doi 10.1109/TAFFC.2019.2895141, 2019.
[14] J. Guo, Z. Lei, J. Wan, E. avots, N. hajarolasv, B. knyazev, A. et. al, "Dominant and Complementary Emotion Recognition from Still Images of Faces", IEEE Access 6, doi 10.1109/ACCESS.2018.2831927, 2018, pp. 26391-26403.
[15] Adeyanju IA, Omidiora EO, Oyedokun OF (2015) Performance evaluation of different support vector machine kernels for face emotion recognition. In: SAI Intelligent Systems Conference (IntelliSys), IEEE, 10–11 November 2015, London, UK
[16] Bakariya B, Thakur GS (2015) An efficient algorithm for extracting high utility item sets from weblog data. IETE Tech Rev 32(2):151–160
[17] Benamara NK et al (2021) Real-time facial expression recognition using smoothed deep neural network ensemble. Integrat Comput-Aided Eng. https://doi.org/10.3233/ICA-200643
[18] Boragule A, Akram H, Kim J, Jeon M (2022) Learning to resolve uncertainties for large-scale face recognition. Pattern Recogn Lett 160:58–65



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)