iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VI **Month of publication:** June 2020

DOI: http://doi.org/10.22214/ijraset.2020.6195

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: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

A Comparative Analysis on Hand Gesture Recognition using Deep Learning

Preeti Nutipalli[1], Surya Pavan Kumar Gudla[2], B.Yogitha[3], Gujjala Rajesh[4]
[1, 2]*Assistant Professor, Department of CSE,* [3, 4]*Final Year B.Tech. Students, Aditya Institute of Technology and Management, Tekkali, A.P, India,*

Abstract: ***In recent years, the vision-based innovation of hand motion acknowledgment is a significant piece of human computer interaction (HCI).In the last decades; keyboard and mouse play a significant role in human-computer interaction. However, owing to the rapid development of hardware and software, new types of HCI methods have been required. In particular technologies, such as speech recognition and gesture recognition receive great attention in the field of HCI. Hand gesture recognition is very significant for human-computer interaction. On survey many models were used to recognize hand gestures with different custom images from various datasets captured by camera. One of the approaches that highly used in image feature extraction is Convolution neural network (CNN). In this work, we present a novel real-time method for hand gesture recognition. In our framework, the hand gesture is detected from the trained dataset of images. Then, the palm and fingers are segmented so as to detect and recognize the fingers. Finally, a rule classifier is applied to predict the labels of hand gestures. The experiments on the data set of 2569 images show that our method performs well and is highly efficient. Finally we have shown the comparison among the CNN with Softmax and KNN classifier on the images which leads accurate result.***
Keywords: HCI, Hand Gestures, Machine Learning, CNN, KNN, Softmax Classifier

I. INTRODUCTION

PCs have become a key component of our general public since their first appearance. Every literate and illiterate able to interact with the computer system with ease, as interfaces came into evolution. Different interfaces were designed based on need. Early Graphical user interface (GUI) designed to interact with system by clicking on icons without need of any internal commands. Later on advancement of technology many interfaces were designed to meet the needs and provide flexibility of users.

Various types of recognition systems are like Facial Recognition System, Speech Recognition System, Pattern Recognition System and Optical Character Recognition System etc.;. Each recognition system has its own model to predict the output for given input. Many algorithms are designed to recognize the required result.

Recently, there has been a surge in interest in recognizing human hand gestures. Hand gesture recognition has various applications like computer games, machinery control (e.g. crane), and thorough mouse replacement.

Computer recognition of hand gestures may provide a more natural-computer interface, allowing people to point, or rotate a CAD model by rotating their hands. Hand gestures can be classified in two categories: static and dynamic. A static gesture is a particular hand configuration and pose, represented by a single image. A dynamic gesture is a moving gesture, represented by a sequence of images. Interactive applications pose particular challenges. The response time should be very fast. The user should sense no appreciable delay between when he or she makes a gesture or motion and when the computer responds. The computer vision algorithms should be reliable and work for different people. Generally poses and gestures are one the basic means of communication between humans and system as they are able to transfer some kind of meaning. The examination region of posture furthermore, motion acknowledgment expects to perceive such articulations, which regularly include some stance or potentially movement of the hands, arms, head, or even skeletal joints of the entire body. In gesture recognition, there are certain image processing features, which are trailed by neural systems to help order the distinguished motion. Some of them are i) Image Preprocessing ii) Segmentation iii) Image Enhancement iv) Morphological filtering.

1) *Image Preprocessing:* Captured images are preprocessed using two modes. a) Binary Mode b) Skin Mask Mode. Further noise removal techniques like Gaussian blur erosion and dilation methods are applied in each of these modes.
2) *Segmentation:* Image Segmentation is a basic procedure in PC vision. It includes partitioning a visual contribution to portions to disentangle image examination. Here skin segmentation is used in order to reduce from non skin back ground as people have different complexions, in order to achieve optimal threshold value by using adaptive thresholding process.
3) *Image Enhancement:* It is used in order to improve brightness and reduce blurring in images while image acquisition by using Gaussian blur.

4) *Morphological Filtering:* In order to get a better smooth, closed and contour of gestures, on segmented images morphological filtering is applied. This can be done by erosion and dilation operation.

Here in the application compared convolution neural network with KNN a class of deep neural network which uses softmax classifier of CNN, which is highly used for analyzing visual images and leads to accurate results.

In this paper, we present a motion acknowledgment approach that centers on hand motions. A gesture is a non verbal communication in which visible body communicates particular message. Various hand gestures are examined extracted from trained dataset, and allowed them to classify to recognize exact action by CNN model.

The rest of the paper is organized as follows: section II presents presents related research works in the territory of motion acknowledgment utilizing skeletal information, concentrating on those that are in view of profound learning. Section III presents proposed work of how CNN model used to extract hand gestures on dataset containing trained images. Section IV presents Experimental Results and comparative analysis of CNN with Softmax and CNN with KNN. Finally in section V we present our conclusion.

II. RELATED WORK

The concept of hand gesture recognition is highly attracted by many authors and made to design various models and algorithms for identification of gestures. In this paper we identified and listed out different algorithms used and proposed by many authors. And finally we used machine learning algorithms and made comparison with classifiers which presents high accuracy.

Eirini Mathe et al presented an approach towards hand gesture recognition that uses a Convolutional Neural Network (CNN), which is trained on Discrete Fourier Transform (DFT) images that result from raw sensor readings. From a given set of video frames we obtain the 3D coordinates of a set of skeletal joints. These are concatenated to form a signal image which is then transformed to an activity image using the Discrete Fourier Transform. The proposed Convolutional Neural Network is trained using these activity images.

Simran Shah et al used Convolutional Neural Networks. It is a class of deep neural networks which is most accurate and efficiently applied for analyzing visual imagery. Convolutional Neural Networks use a slight variation of multilayer perceptrons. This is designed so that it required minimal preprocessing, resulting in better experiment accuracy[1].

Rutuja J et al used Convolutional Neural Networks (CNN) over data set has total 27456 entries and 784 labels for training, out of which 30% are used for validation. We are randomly splitting the dataset by subject into training 80% and testing 20%. All the images were stored as per their pixel values after they were converted into a grey scale image and able to get an accuracy of 95.6% which is higher than all the previously implemented systems[2].

K. Narmatha et al a deep convolution neural network is proposed to immediately classify hand gestures in pictures without any segmentation or detection stage that could discard the irrelevant not-hand area**.** For each image, human face and human body are the main components of the image, while the hand gesture to be classified only occupies the 10% of pixels the entire image. To get a reliable result, people who appear in the testing set are totally distinct from the training set. Cross-validation strategies are adopted, everyone with 25 people for training, 5 person for validation, and 10 people for testing [3].

From all readings of various papers [1][6][7] we focus on classification of images and indentify the corresponding gesture. Here we considered five gestures which respond as OK, PEACE, PUNCH, STOP and NOTHING

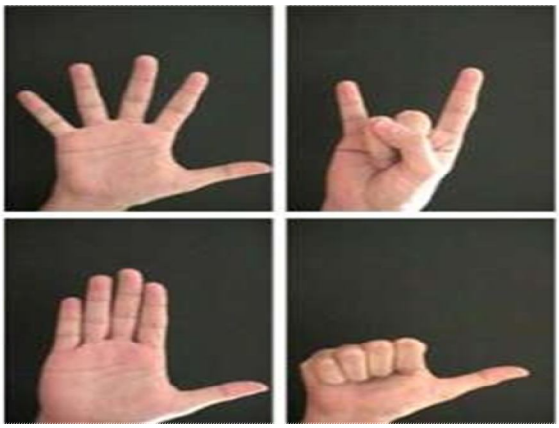


Figure 1: Few Hand Gestures

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

III. PROPOSED WORK

With regard to this work, the application provides functionalities like Prediction, Training and Visualization. The picture database can have various arrangements. Pictures can be either hand drawn, digitized photos or a 3D dimensional hand. Photos were utilized, as they are the most practical methodology. In the principal push there are the preparation pictures. In the second there are the trying pictures.

Figure 2: Training Images

Figure 3: Testing Images

We present a signal acknowledgment approach that centers on hand motions. We propose a novel profound learning engineering that utilizes a Convolutional Neural Network (CNN).

All the more explicitly, we utilize the Kinect sensor and its Software Development Kit (SDK) so as to distinguish and follow the subject's skeletal joints in the 3D space[4]. We at that point select a subset of these joints, i.e., all that are included at any of the signals of our informational collection. At that point, we make a counterfeit picture dependent on this 3D arranges. We apply the Discrete Fourier Transform on these pictures and utilize the subsequent ones to prepare the CNN.

We contrast our methodology and past work, where a lot of hand-created factual highlights on joint directions had been utilized. At long last, we exhibit that it is conceivable to productively perceive hand motions without the need of a component extraction step. Assessment happens utilizing another dataset of 10 hand motions.

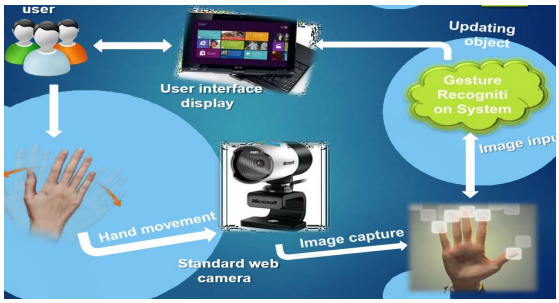


Figure 4: Basic working of Gesture Working Technology

In our venture we have utilized convolutional neural systems. It is a class of profound neural systems which is generally exact and effectively applied for investigating visual symbolism. Convolutional Neural Networks utilize a slight variety of multilayer perceptron. This is planned with the goal that it required negligible preprocessing, bringing about better examination precision. They are additionally called move invariant or space invariant fake neural systems (SIANN), because of their interpretation invariance attributes and shared-loads engineering. These neural systems attract their fundamental standard from the way which the creature visual cortex is composed.

There are numerous availability designs among neurons, and the neural systems are propelled by this. There is a confined locale of the visual field which is known as the responsive field. The individual cortical neurons react to upgrades just in this field. These open fields of various neurons halfway cover in a way with the end goal that they spread the whole visual field.

A. Convolutional Neural Network

Convolutional Layer is the primary layer to extricate highlights from an info picture. Convolution assists with safeguarding the connection between different pixels by learning picture highlights utilizing little squares of information. It is a numerical activity which for the most part takes two information sources. They are the picture network and a channel or a bit [1][5].

Stride: This is the quantity of pixels moves that we perform over the information framework. For instance, when the step is 1 then we move the channels to 1 pixel at once. Essentially, when the step is 2 then we move the channels to 2 pixels one after another, etc.

Padding: It is seen that on occasion the channel doesn't fit flawlessly fit the information picture. We have two primary alternatives for this situation: a) Pad the image with zeros (zero-cushioning) so it fits or b) drop the piece of the picture where the channel didn't fit. The subsequent technique is called substantial cushioning which keeps just legitimate piece of the picture

Non-Linearity: ReLU implies Rectified Linear Unit for a non-straight activity. The yield is f(x) = max (0, x). ReLU is very significant in CNNs: the primary object is to present non-linearity in our ConvNet. Since, this present reality information would need our ConvNet to realize what could be non-negative direct qualities.

Polling: This area would decrease the quantity of parameters when the pictures are excessively enormous. Spatial pooling likewise called down inspecting or sub sampling in light of the fact that it diminishes the dimensionality of each guide however holds the significant data. Spatial pooling can be of various sorts a) whole pooling b) max pooling c) normal pooling

Fully Connected Layer: In this last layer, we smooth our grid into a vector and feed it to a completely associated layer like the proposed neural system.

- An image matrix (volume) of dimension **(h x w x d)**
- A filter **(f_h x f_w x d)**
- Outputs a volume dimension **(h - f_h + 1) x (w - f_w + 1) x 1**

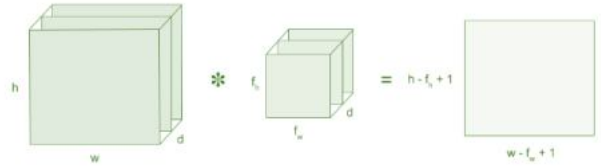


Figure 5: Convolutional layer of CNN

B. Architecture of Proposed System

We have utilized successive API to make our model layer-by-layer.

Our ConvNet for hand signal acknowledgment has the design [INPUT - CONV - RELU - CONV - RELU - MAXPOOL - DROPOUT - FLATTEN - DENSE - RELU - DROPOUT - FC - SOFTMAX]

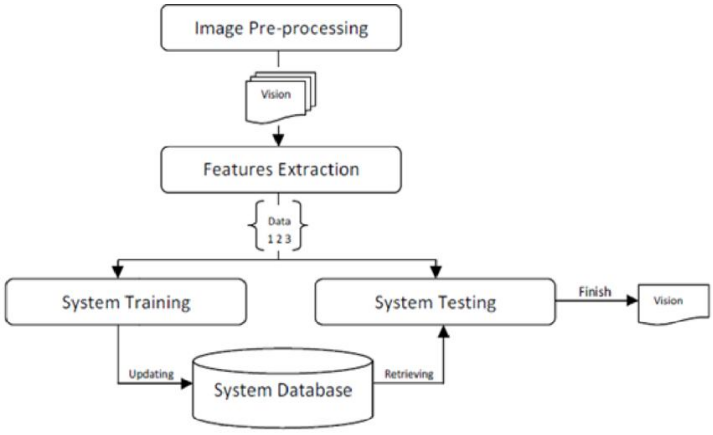


Figure 6: Architecture of System

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

INPUT [200x200x1] will hold the crude pixel estimations of the picture, for this situation a picture of width 200, stature 200, and with 1 shading channel. (dim scale) CONV layer will compute spot item between their loads and a little area they are associated with in the info volume.
The RELU layer will apply an element wise initiation work, for example, the actuation capacity of max (0, x) thresholding at zero.
The POOL layer will play out a subsampling activity along the spatial measurements (width, tallness).
FC (for example completely associated) layer will register the class scores, bringing about volume of size [1x1xnb_classes], where each of the nb_classes no. of numbers relate to a class score. Likewise with other traditional Neural Networks and as the name infers, every neuron in this layer will be associated with all the numbers in the past volume[1].
The target work that the model attempts to limit is unmitigated cross-entropy. The model uses 'adadelta' analyzer.

Layer (type)	Output Shape
conv2d_1 (Conv2D)	(None, 32, 198, 198)
activation_1 (Activation)	(None, 32, 198, 198)
conv2d_2 (Conv2D)	(None, 32, 196, 196)
activation_2 (Activation)	(None, 32, 196, 196)
max_pooling2d_1 (MaxPooling2	(None, 32, 98, 98)
dropout_1 (Dropout)	(None, 32, 98, 98)
flatten_1 (Flatten)	(None, 307328)
dense_1 (Dense)	(None, 128)
activation_3 (Activation)	(None, 128)
dropout_2 (Dropout)	(None, 128)
dense_2 (Dense)	(None, 5)
activation_4 (Activation)	(None, 5)

Figure 7: Layers Summary

S.no.	Architecture Details
1.	Input Image size: (200,200,1)
2.	Total Number of Layers: 12 Total Conv2D layers used: 2 Total MaxPooling2D layers used: 1 Total Fully Connected layers used: 3 Activation layers:4 Dropout layers:2
3.	Kernel size at each Conv2D layer: 3 X 3
4.	Pool size at each maxPooling2D layer: (2,2)
5.	Output Class Labels: 5

Table 1: Architecture Details of CNN

C. Implementation Details

The product and equipment setups used to build up the application are done on
Windows 10 Operating System with Python form 3.6 on manager Anaconda Navigator Spyder. Required equipment arrangements are intel center i5 8th Gen Processor with NVIDIA GEFFORCE GPU on 4GB RAM and 500GB Hard Disk.
The libraries utilized past to create and code the application Open CV 4.3.0, tkinter GUI 8.0, NumPy 1.18.1, Pandas 1.0.3 and PIL 1.1.7

Using the fixed set of images, we have evaluated models like CNN and KNN with classifiers in order to predict gestures which could result high accuracy with no loss. Classifier Softmax applied on CNN classifies with 96.03% accuracy but on comparison with KNN leads to 96.27% accuracy. Here CNN uses by default classifier Softmax, but KNN itself acts as neural network model and uses K-Nearest Neighbor as classifier on images extracted from completion of activation function.

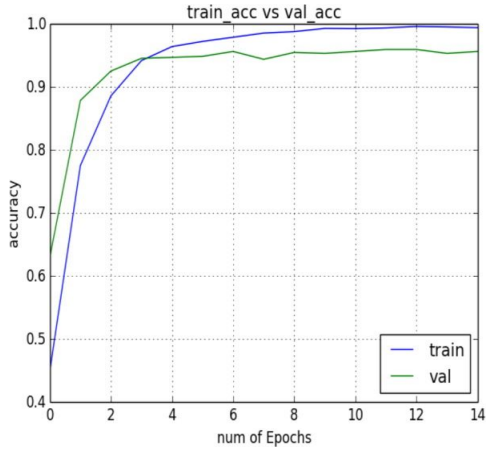


Figure 8: Graph of accuracy against number of epochs Using CNN

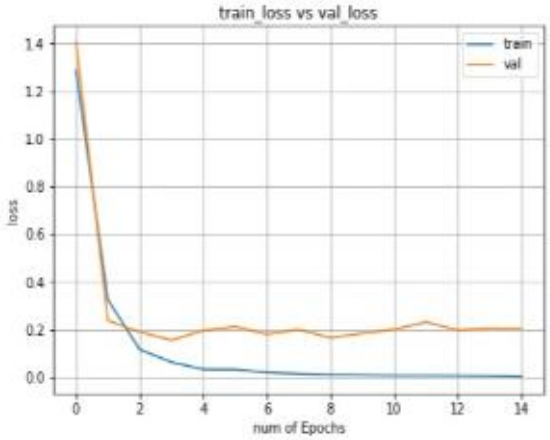


Figure 9: Graph of loss against number of epochs Using CNN

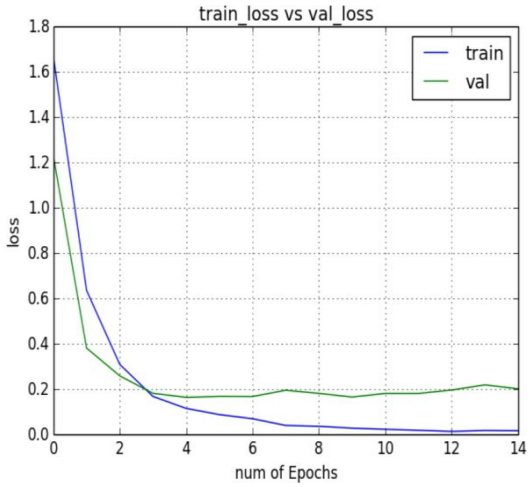


Figure 10: Graph of loss against number of epochs Using KNN

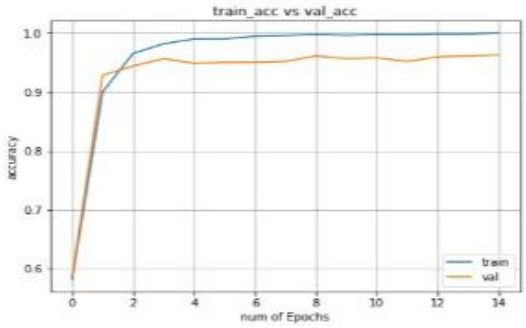


Figure 11: Graph of accuracy against number of epochs Using KNN

D. Comparison Table

Model Used	Training accuracy	Validation Accuracy	Training loss	Validation loss
CNN+Softmax	98.98%	96.03%	0.1048	0.2001
KNN	100%	96.27%	0.0021	0.2027

IV. CONCLUSION

OpenCV is utilized for catching the client's hand signals. Post preparing is performed on the caught pictures to feature the shapes and edges, such as applying parallel limit, obscuring, dark scaling. The Accuracy of the model depends on numerous viewpoints in our dataset and furthermore the highlights present in the preparation information. The dataset was made with no commotion i.e., the motions introduced are sensibly unmistakable; the pictures are clear and without foundation. Additionally there was sufficient number of tests which made our model powerful.

REFERENCES

[1] Simran Shah, Ami Kotia, Kausha Nisar, Aneri Udeshi, Prof. Pramila. M. Chawan, 2019, April. A Vision Based Hand Gesture Recognition System using Convolutional Neural Networks, International Research Journal of Engineering and Technology (IRJET), pages: 2570-257
[2] Rutuja J, Aishwarya J, Samarth S , Sulaxmi R, Mrunalinee P. 2019 April. Hand Gesture Recognition System Using Convolutional Neural Networks, International Research Journal of Engineering and Technology (IRJET), pages: 4508-4514
[3] Kundan Kumar Dubey, Ajitanshu Jha, Akshay Tiwari, K. Narmatha. 2019 April. Hand Gesture Movement Recognition System Using Convolution Neural Network Algorithm International Research Journal of Computer Science (IRJCS), pages: 154-160
[4] Real-Time Hand Gesture Detection and Recognition Using Simple Heuristic Rules- Maria Abastillas, 17 June 2011, Pages:1-57
[5] Shweta. K. Yewale and Pankaj. K. Bharne, "Hand Gesture Recognition Using Different Algorithms Based on Artificial Neural Network", IEEE International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016, pages 671-675.
[6] Hamid A. Jalab, Herman. K. Omer, "Human Computer interface using Hand Gesture Recognition based on neural network", IEEE 06 August 2015 , Electronic ISBN: 978-1-4799-7626-3.
[7] Rafiqul Zaman Khan, Noor Adnan Ibraheem, 2012, July. Hand Gesture Recognition: A Literature Review, International Journal of Artificial Intelligence & Applications (IJAIA), pages: 161-174

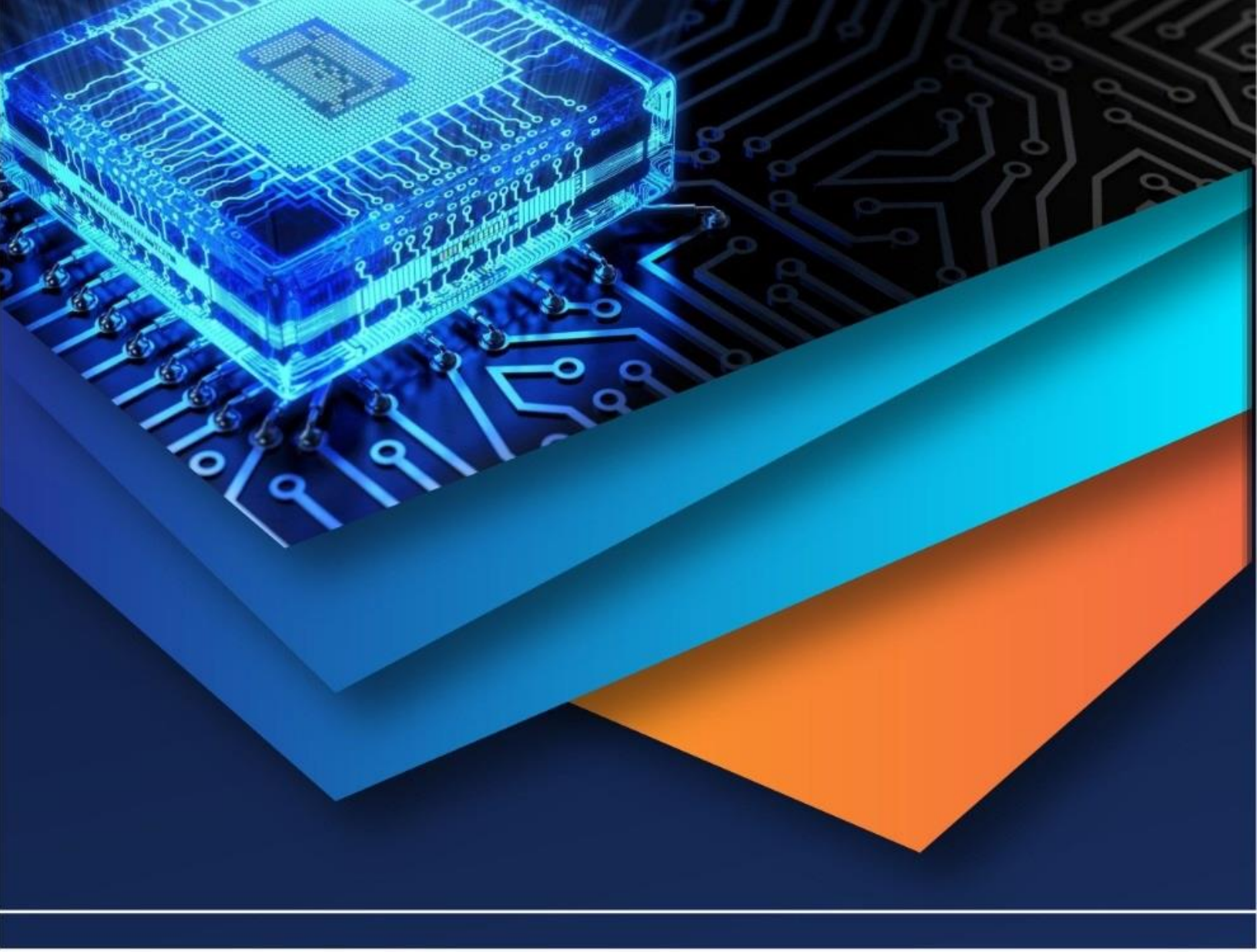
BIOGRAPHIES

A. Ms. Preeti Nutipalli working as Assistant Professor, Dept. of CSE, Aditya Institute of Technology and Management, Tekkali. She had total experience of 11 years in Teaching. Her interested areas Computer Networks, Mobile Computing, Machine Learning.

B. Mr. Surya Pavan Kumar Gudla working as Assistant Professor, Dept. of CSE, Aditya Institute of Technology and Management, Tekkali. He had total experience of 9 years in Teaching. His interested areas are computer Networks, Fog Computing, Machine Learning.

C. Ms.Yogitha Battula student of final year, CSE of Aditya Institute of Technology and Management, Tekkali. Done internship on Project titled "Power Utility - New Service Connection Module" at Fluentgrid Limited. Got placed in HCL Technologies, presently working in MAQ Software.

D. Mr.Gujjala Rajesh student of final year, CSE of Aditya Institute of Technology and Management, Tekkali. Got placed in Effectronics, Vijayawada.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY