iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38058

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 9 Issue IX Sep 2021- Available at www.ijraset.com

Sign Language Recognition using CNN

Aniket Wattamwar[1], Jay Shiurkar[2], Pooja Lakhane[3], Shilpa Lambor[4]
[1, 2, 3, 4]*Electronics Engineering Department, Vishwakarma Institute of Technology, Pune*

Abstract: ***This research work presents a prototype system that helps to recognize hand gesture to normal people in order to communicate more effectively with the special people. Aforesaid research work focuses on the problem of gesture recognition in real time that sign language used by the community of deaf people. The problem addressed is based on Digital Image Processing using CNN (Convolutional Neural Networks), Skin Detection and Image Segmentation techniques. This system recognizes gestures of ASL (American Sign Language) including the alphabet and a subset of its words.***
Keywords: gesture recognition, digital image processing, CNN (Convolutional Neural Networks), image segmentation, ASL (American Sign Language), alphabet

I. INTRODUCTION

Technology in the last decade have become an important aspect of life and are used in various fields however, the systems and methods that we use to interact with computers are outdated and have various issues. Although, computers have made plentiful advancement in both fields of Software and Hardware, still the fundamental way in which Humans try to communicate with technological product is using basic pointing device like a mouse and Keyboard or advanced Voice Recognition System, or maybe Natural Language processing in really advanced cases to make this communication more human and easy for us. Our proposed method in this project is the Hand gestures recognition system to replace the basic pointing devices used in computer systems. This will definitely bridle the use of equipment like mouse and touchpad. Gesture is a form of non-verbal information. A person can communicate with other people through gestures and signs. The camera acts as the eyes of the computer to perceive the gestures made by the human. A very popular research among researchers working in the domain of computer vision. Everyone has access to a laptop or a mobile phone nowadays which has a camera inbuilt. Using this camera, we can bring about a natural way of recognizing these gestures for the people who are most like to use it. Since it is accessible to almost anyone creating an app out of this application seems to be the most appropriate and beneficial way. Moreover, as virtual reality devices becoming more common like VR Goggles, the laptop or the phone camera may also become just an interaction between the real world and digital one. This paper looks at solving the problem of recognizing the gestures through the camera of the laptop. This proof of concept can work as the base of creating a mobile app in the future. Keeping the same free for the users to download them and use them. The app can be used by pointing the camera to the person who is deaf and dumb and can only interact through Sign Language. The method proposed by us uses the concept of Neural Networks on images, in this case, the image being the hand sign made by the user.

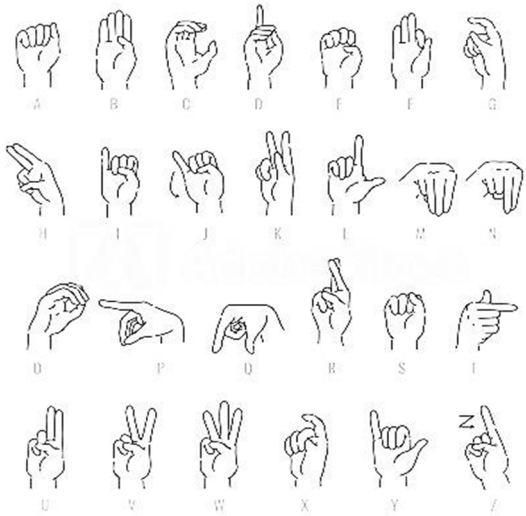


Figure 1 – ASL
Source: Google

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

II. RELATED WORK

[1] implements the recognition by finding the orientation of the gesture. It also checks the thumb of the hand sign and the finger region. This is done to check the peaks of the fingers. Once all the peaks are generated using the Euclidean distance, count of how many fingers are raised can be calculated. Elmahgiubi [2] in his survey paper 'Gesture Recognition- A Survey' talks about other methods of implementing Gesture Recognitions such as Hidden Markov Models (HMM's), Finite State Machines (FSM), Neural Networks. The neural networks being adaptive and robust in nature are highly chosen as recognition system algorithms. Further describing about the recognitions of face and head using HMM's

Lu, W.,Tong, Z., & Chu, J. [4] in their paper of 'Dynamic Hand Gesture Recognition with Leap Motion Controller' says about different sensors used for real-time recognition of hand movements. The LMC, a depth sensor is developed in a way which gives three-dimensional depth data of the scene. The features in their paper include palm direction, fingertip orientation, palm center position. For classification they have used HCNF classifier post feature extraction. A similar kind of approach using depth sensor is given by [6] Suarez, J., & Murphy, R. R in their review paper.

M., Kallel, M., & Bouhlel [7] discuss about applications of gesture recognition. Further, algorithms like Otsu's method for thresholding are used, skin detection, hand movement tracking. Skeleton method is also described in the paper which uses static and dynamic signatures for recognition.

Beke, A., Yuceler [11] shows an incredible way of gesture recognition using Fuzzy Logic. They have integrated with their smart phone. It imitates the response of the gesture using the IMU sensor data. With the help of a rule-based fuzzy classifier is processed.

III. PROPOSED SYSTEM

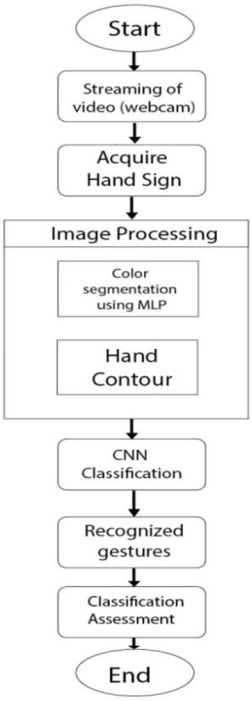


Figure 2 – Flowchart

We propose a method to identify the gestures image processing, classification techniques. Figure 2 depicts the complete flow of the process. This process will provide optimum results. However, accuracy can be achieved by introducing various complex systems. We obtain the video stream through webcam. Then the sign displayed is captured and further an image is acquired for processing. The image processing part is the vital for detection of the accurate sign. The color of the image that is the part of the hand is highlighted by the multilayer perceptron image segmentation technique. Outlines or boundaries of the hand sign are plotted which are nothing but the contours. Then the classification of the sign is done using CNN algorithm. The input test image is taken and the entire preprocessing is done on that image too. This input image is used and cross correlation is calculated with all the images in the database.

©IJRASET: All Rights are Reserved

A. Convolutional Neural Network (CNN)

The next part of the video analysis for tracking and creating mask is the CNN. As the CNN consists of inputs and output layers, we need to find a way of. converting the image into layers of input containing values. The input image is convolved with a filter or a kernel matrix and is called a feature map. The output of this is a convolved image. CNNs are used for classification, recognition, Video Analysis, NLP and more.

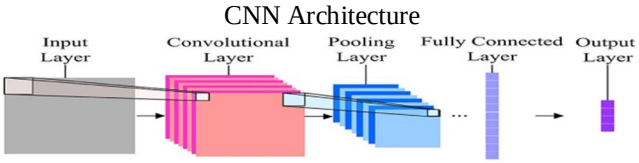


Figure 3 – The CNN Architecture

Understanding the CNN from the above figure. The input layer is our image with pixel values. This is how it works: Convolution Layer, Pooling, Fully Connected Layer, Output Layer.

1) *Convolutional Layer:* This is the first layer to extract features from an input image with the help of a stride. It is a matrix multiplication operation that takes two inputs such as image matrix and a kernel.
2) *Pooling:* Pooling layer section is done to reduce the size of the image to some extent. Spatial pooling also called subsampling performs dimensionality reduction but does not remove the information from the image. Spatial pooling can be of different types:

a) Max Pooling
b) Average Pooling
c) Sum Pooling.

Max pooling takes the element with the largest value from the feature map. Taking the highest value element could also take the average pooling. Sum of all values in the feature map is called as sum pooling.

3) *Fully Connected Layer:* This layer, we flatten our matrix into vector and is given as input to the fully connected layer like neural network. This layer formed after we have done the flattening of the image. Flattening is basically converting m x n matrix into a single layer.
4) *Output Layer:* This layer contains the useful information taken from the images. This is also now the input layer of the artificial neural network (ANN). Using different activation function, we can solve the ANN.

B. Contours

Contours play an effective role in understanding the shape of an image. It solves the problem of edge detection. As edges have no identity, no continuity and they cannot be differentiated. So, it becomes arduous to work with them. Contours are the solution. Contours are basically the outline of a shape that we have in the image. They are used in object recognition and shape matching. With the help of Contours, area is calculated enclosed in the shape. Contours are part of the pre-processing. Once the skin is detected we can use contours to get shape of the hand sign. This is ensuring that the correlation is calculated only of the hand sign and not the unwanted part. Figure shows the contour on one image of the database. As shown, the entire shape of the hand sign is covered and none of the unwanted part.

C. Image Segmentation

Image segmentation is an essential process in computer vision or Image Processing. It breaks down an input image into segments to simplify image analysis. Segments or parts of objects, amount to sets of pixels, called as “super-pixels”. Image segmentation rearranges pixels into larger components, eliminating the requirement to consider single pixels as single units of observation. There are three levels of image analysis:

1) *Classification:* This type classifies the whole image into a class such as “fruit”, “tree”, etc.
2) *Object Detection:* It recognizes a particular entity in an image and while doing so creates a bounding box around it for identification.
3) *Segmentation:* It recognizes the segment or part of the object and tries to relate to which class it may belong to. Segmentation is the fundamental step for performing object detection and classification.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

D. Skin Detection

The method used in Image Processing of locating human skin-colored pixel values and regions in an image or a video is Skin Detection. This is a pre-processing step to find regions that possibly have human faces and other body parts in images. Several computer visions approaches have been developed for skin detection. A detector converts a given pixel value in the image into an appropriate color range and then use a classifier to classify the pixel into skin or otherwise. A skin classifier defines a decision boundary like contours around the skin color class based on a training database of skin-colored pixels.

IV. RESULTS

As mentioned the input test image is compared with the ones in the database to give the similarity match. Figure 6 shows the array with the corresponding class/alphabet it belongs to.

For this paper, we have used 5000 images with 1000 images belonging to one group/class. The model is trained on these 5000 images. The validation test set contains 100 images per class. All the results are based on the above data. The accuracy of the model can be increased by increasing the number of images given to the model during training.

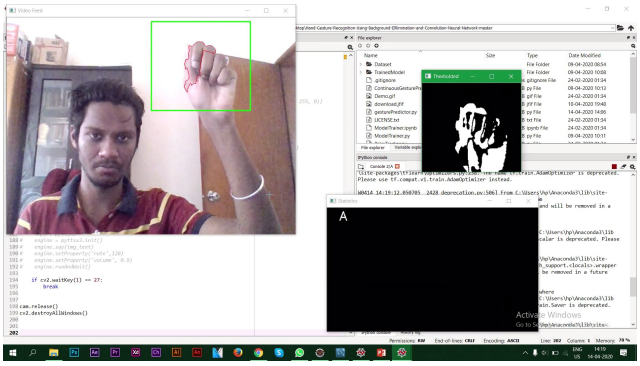


Figure 4 – Letter ‘A’ from the sign

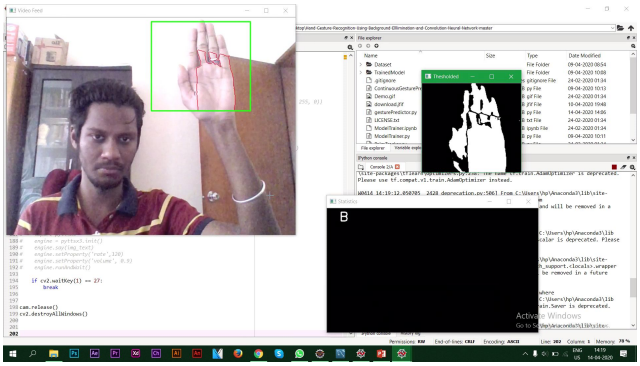


Figure 5 – Letter ‘B’ from the sign

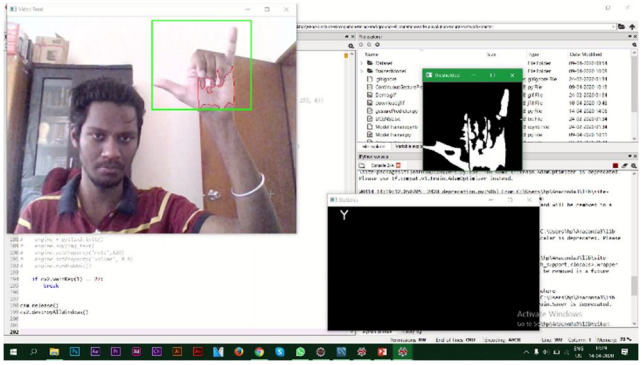


Figure 6 – Letter ‘Y’ from the sign

©IJRASET: All Rights are Reserved

V. CONCLUSION

With the help of this technology it will be easy to communicate with dumb people with faster pace. Also, people can communicate and understand the message that dumb people wants to convey efficiently. Complete sentences and paragraphs can be formed and can be used for various applications. A real time Application for mobile users can be developed so that most of the people can access the system easily.

With the use of complex algorithm such as Neural Networks, a better model can be made for real time recognition. The output can further be converted to speech for better User Experience

REFERENCES

[1] Panwar, M. (2012). Hand gesture recognition based on shape parameters. 2012 International Conference on Computing, Communication and Applications. doi:10.1109/iccca.2012.6179213 J. Dai, K. He, and J. Sun. Convolutional feature masking for joint object and stuff segmentation. In CVPR, 2015.

[2] Mitra, S., & Acharya, T. (2007). Gesture Recognition: A Survey. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 37(3), 311–324. doi:10.1109/tsmcc.2007.893280

[3] Nikam, A. S., & Ambekar, A. G. (2016). Sign language recognition using image based hand gesture recognition techniques. 2016 Online International Conference on Green Engineering and Technologies (IC-GET). doi:10.1109/get.2016.7916786

[4] Lu, W., Tong, Z., & Chu, J. (2016). Dynamic Hand Gesture Recognition With Leap Motion Controller. IEEE Signal Processing Letters, 23(9), 1188–1192. doi:10.1109/lsp.2016.2590470

[5] Elmahgiubi, M., Ennajar, M., Drawil, N., & Elbuni, M. S. (2015). Sign language translator and gesture recognition. 2015 Global Summit on Computer & Information Technology (GSCIT). doi:10.1109/gscit.2015.7353332

[6] Suarez, J., & Murphy, R. R. (2012). Hand gesture recognition with depth images: A review. 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. doi:10.1109/roman.2012.6343787

[7] Ben Abdallah, M., Kallel, M., & Bouhlel, M. S. (2012). An overview of gesture recognition. 2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT). doi:10.1109/setit.2012.6481883

[8] Borg, M., & Camilleri, K. P. (2019). Sign Language Detection “in the Wild” with Recurrent Neural Networks. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp.2019.8683257

[9] Malgireddy, M. R., Corso, J. J., Setlur, S., Govindaraju, V., & Mandalapu, D. (2010). A Framework for Hand Gesture Recognition and Spotting Using Sub-gesture Modeling. 2010 20th International Conference on Pattern Recognition. doi:10.1109/icpr.2010.921

[10] Lee, D., Yoon, H., & Kim, J. (2016). Continuous gesture recognition by using gesture spotting. 2016 16th International Conference on Control, Automation and Systems (ICCAS). doi:10.1109/iccas.2016.7832502

[11] Beke, A., Yuceler, A. A., & Kumbasar, T. (2017). A rule based fuzzy gesture recognition system to interact with Sphero 2.0 using a smart phone. 2017 International Artificial Intelligence and Data Processing Symposium (IDAP). doi:10.1109/idap.2017.8090191

[12] Kar, A., & Chatterjee, P. S. (2015). An Approach for Minimizing the Time Taken by Video Processing for Translating Sign Language to Simple Sentence in English. 2015 International Conference on Computational Intelligence and Networks. doi:10.1109/cine.2015.40

[13] T. Starner and A. Pentland, “Real-time American Sign Language recognition from video using hidden Markov models,” MIT Media Lab., MIT, Cambridge, MA, Tech. Rep. TR-375, 1995

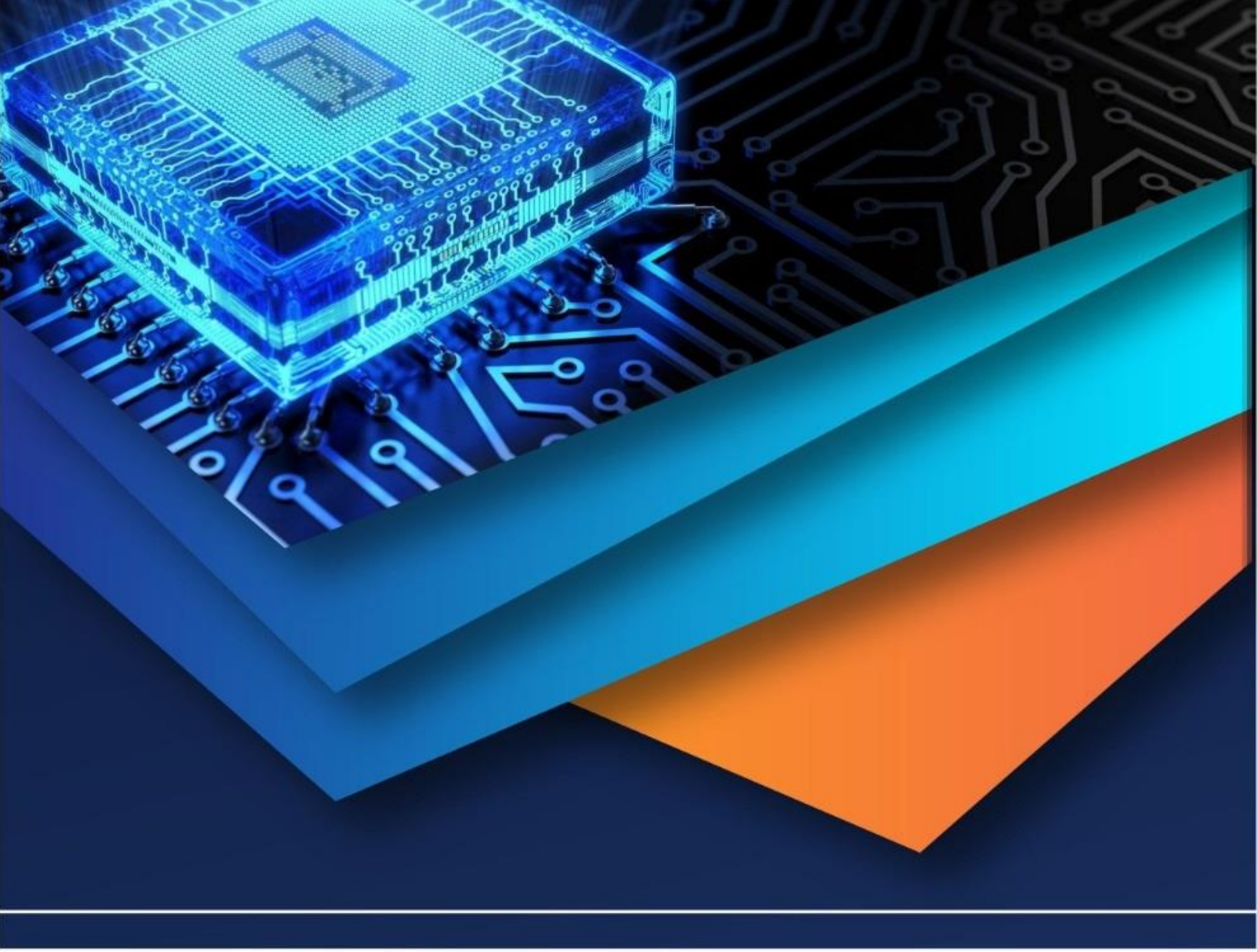
[14] H. S. Yoon, J. Soh, Y. J. Bae, and H. S. Yang, “Hand gesture recognition using combined features of location, angle and velocity,” Pattern Recogn., vol. 34, pp. 1491–1501, 2001.

[15] A. Ramamoorthy, N. Vaswani, S. Chaudhury, and S. Banerjee, “Recognition of dynamic hand gestures,” Pattern Recogn., vol. 36, pp. 2069–2081, 2003.

[16] J. Davis and M. Shah, “Visual gesture recognition,” Vis., Image Signal Process., vol. 141, pp. 101–106, 1994

[17] J. Yamato, J. Ohya, and K. Ishii, “Recognizing human action in timesequential images using hidden Markov model,” in Proc. IEEE Int. Conf.Comput. Vis. Pattern Recogn., Champaign, IL, 1992, pp. 379–385

©IJRASET: All Rights are Reserved



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)