

# Design of an Intelligent Approach on Capsule Networks to Detect Forged Images

**J. Samuel Manoharan**

Professor, Department of Electronics and Communication Engineering, Sir Isaac Newton College of Engineering and Technology, Nagapattinam, India

**E-mail:** drjasm1530@ieee.org

## Abstract

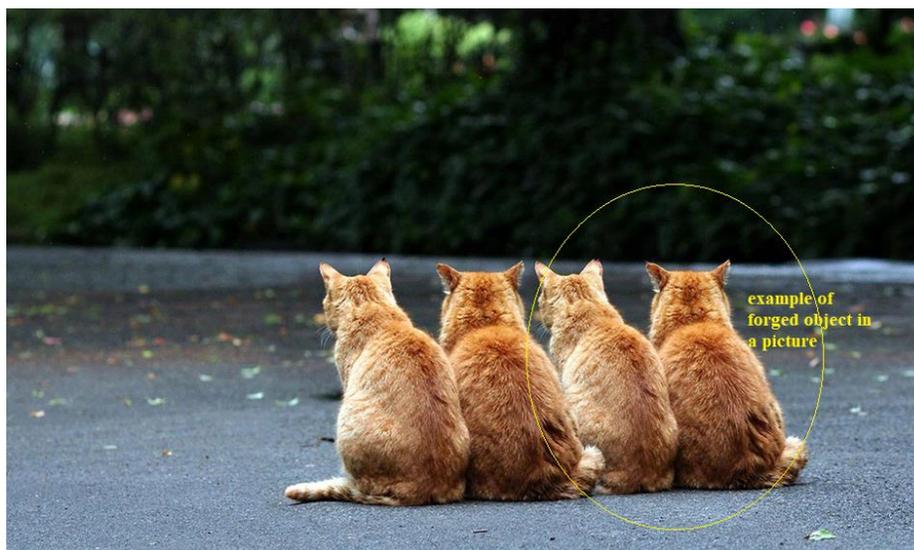
Forgeries have recently become more prevalent in the society as a result of recent improvements in media generation technologies. In real-time, modern technology allows for the creation of a forged version of a single image obtained from a social network. Forgery detection algorithms have been created for a variety of areas; however they quickly become obsolete as new attack types exist. This paper presents a unique image forgery detection strategy based on deep learning algorithms. The proposed approach employs a convolutional neural network (CNN) to produce histogram representations from input RGB color images, which are then utilized to detect image forgeries. With the image separation method and copy-move detection applications in mind, the proposed CNN is combined with an intelligent approach and histogram mapping. It is used to detect fake or true images at the initial stage of our proposed work. Besides, it is specially designed for performing feature extraction in image layer separation with the help of CNN model. To capture both geographical and histogram information and the likelihood of presence at the same time, we use vectors in our dynamic capsule networks to detect the forgery kernels from reference images. The proposed research work integrates the intelligence with a feature engineering approach in an efficient manner. They are well-known and efficient in the identification of forged

images. The performance metrics such as accuracy, recall, precision, and half total error rate (HTER) are computed and tabulated with the graph plot.

**Keywords:** Capsule Networks, Forged Image Detection

## 1. Introduction

Due to the recent advancements in digital image processing technology and the widespread usage of digital cameras, editing or tampering with a digital image has never been simpler. Even inexperienced forgers may now readily modify digital photos using simple photo editing tools like Adobe Photoshop. Drooled photos have become increasingly common and sophisticated in the last few decades with a seemingly endless stream of digital forgery tools has appeared, among which splicing and copy-move are the most common ones that manipulate the images in a way that they are hardly perceived by the human perceptual system that has appeared previously [1-5]. As a result, digital image forensics places a high priority on effectively detecting these two types of forgeries. Figure 1 example of forged image.



**Figure 1.** Example of Forged Image

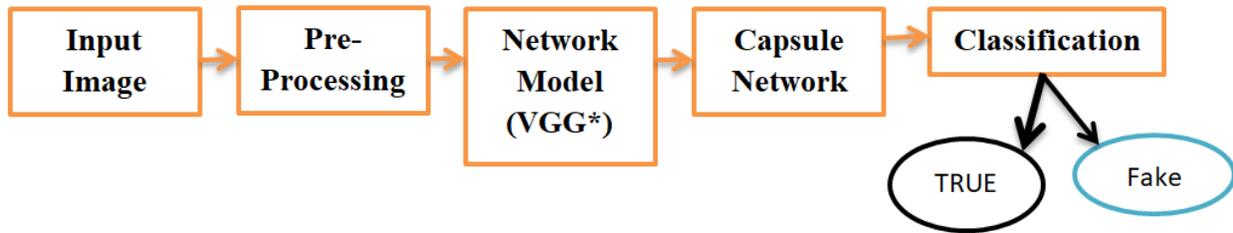
Physiological features or behavioral aspects of people are used by biometric recognition methods to identify a specific person. False pictures and videos may be created using forged photographs and videos to get beyond face authentication and spread fake news. Using huge quantities of training data and sophisticated network topologies, the quality of altered pictures and videos has significantly improved [6-11]. As a result, the process of creating fake faces has been made simpler. A brief video or an ID picture is all that's required these days to generate a fabricated face appearance. Figure 2 shows the example of authentic and tampered or sliced based forged image.



**Figure 2.** Example of Authentic and Tampered / Sliced based Forged Image

Social networks are becoming important an information source, which means that altered multimedia, may rapidly spread and have substantial impacts. This danger is well-illustrated by the deep fake phenomena. A human image synthesis method based on artificial intelligence allows

anybody with access to a computer to make movies, including the appearance of any celebrity [12-14]. Figure 3 shows the basic block diagram of capsule network structure.



**Figure 3.** Basic Capsule Network Structure

Using vectors and the likelihood of existence at the same time, capsule networks are new deep learning models that collect spatial information. Detecting faces, for example, requires the knowledge of features such as eye and nose positions. Different speech processing applications, including command and emotion detection have been investigated using capsule networks. Replay assaults employing printed pictures or recorded films on computer-generated movies have recently been utilized for the identification of forgeries produced from forged images and videos [15-17]. In terms of detecting such assaults, research shows that capsule networks outperform other cutting-edge systems.

## 2. Organization of the Research

The rest of this research paper is structured as follows: Section 3 describes the current prior detection technique for forged images; Section 4 describes the suggested technique for detecting and classifying counterfeit pictures in an effective manner. Finally, Section 5 summarizes the outcome of the proposed experiment. Section 6 concludes the proposed research work with future research directions.

### 3. Preliminaries

An image splicing detection approach based on natural picture features was presented by Shi et al. Image features extracted from statistical moments of characteristic functions were combined with the Markov transition probability matrices in both the spatial and DCT domains, resulting in discriminative feature vectors for support vector machine (SVM) classification, which were then used in a block discrete cosine transform of the images [18].

Chung et al. and Suwajanakorn et al. developed a technique to assist attackers in understanding the mapping between speech and lip movements, allowing anybody to generate completely synthesized audio-video data [19, 20].

Compared to LBP-based techniques, Kim et al. presented a new method based on local patterns of diffusion speed (local speed patterns). Deep learning has made it possible to identify replay assaults much more easily now [21]. Different methods are used for classifying the features retrieved by pre-trained convolutional neural networks, including Yang et al. [22].

Menotti et al. take a similar approach but improve the filters in a easily accessible high-performance CNN architecture [23]. An additive operator splitting technique is used by Alotaibi and Mahmood in their own CNN as a nonlinear diffusion-based method [24].

Using a 2-D non causal Markov model, Zhao et al. figured out how to localize splicing in images with a high degree of accuracy [25]. The alternative method suggested by Lyu et al. differed significantly from the model-based approaches mentioned above, as it revealed the inconsistencies of local noises caused by camera sensors or post-processing to identify and locate the spliced images [26]. A local binary pattern (LBP) and a steerable pyramid transform (SPT) were used to identify the distortions of textural characteristics in forged images since image tampering may

change the texture micro-patterns. The LBP and SPT have obtained the best detection performance on the CASIA dataset [27].

In 2017, Hinton et al. addressed the shortcomings of CNNs when they are used for inverse graphics applications and provided the groundwork for a more robust capsule design. The absence of an efficient algorithm and computer hardware constraints meant that this sophisticated design could not be executed successfully at the time. Instead, CNN, which is both simple to build and train is now extensively utilized [28]. The dynamic routing algorithm and expectation maximization routing algorithm have now been introduced, and capsule networks have been built with impressive first results. Two recent studies have shown that the hierarchical pose connections between object components may be well represented by the agreement between the capsules computed by the dynamic routing method [29].

## **Research Gap**

These techniques, as well as others must be refined in order to enhance the accuracy of visual tasks. The application of a capsule network to a forensics job, which is the subject of this research paper, is a difficult issue to solve. However, the agreement between capsules obtained via the use of the dynamic routing algorithm may be able to improve detection performance on complicated and almost perfect forged images in certain situations.

## **4. Proposed Methodology**

In general, capsule networks replace CNNs' scalar neurons with high-dimensional vectors instead. The length and direction of the vector are interpreted by representing the likelihood of the entity's existence and instantiation parameters respectively.

#### **4.1 Design of dynamic capsule network**

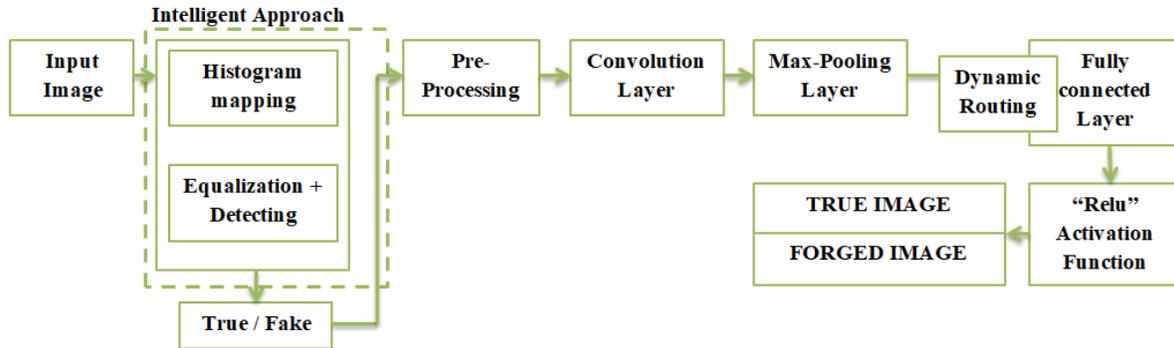
The proposed method generates important histograms by including the values of the whole image's pixels. Certain quantization effects are visible on these histograms only if a picture has been saved in JPEG format several times. If the effect is found, we may deduce that the picture has been altered (or at the very least saved) at least once. This is referred to as the sensible method for initiating the deep learning process.

This research paper proposes a capsule network with dynamic routing. A coordinate picture is formed by our eyes and the brain, and a variation in coordinate frames has a significant impact on how we think. The coordinate frame is thus engaged in the recognition process when the objects are recognised, which is dominated by the notion of space. Thus, instead of representing a single neuron, the capsule depicts an entity by utilizing a network of neurons [30].

Only the images can be implemented by using the new approach. Layers are used for picture input, and the feature engineering process is separated from that. The frames are used to obtain the classification results (posterior probabilities). To arrive at the outcome, the probabilities are averaged in the post-processing stage. The rest of the picture remains as a copy of the original.

#### **4.2 Feature engineering**

Faces (objects) in the given input images are identified and scaled to 128\*128 in the pre-processing step. The latent characteristics are extracted from the VGG-19 network and feed them into the capsule network as input. Instead of three outputs before the ReLU layers, the third max-pooling layer's output [31]. We do this to minimize the size of the capsule network's inputs.



**Figure 4.** Overall Proposed Structure

### 4.3 Algorithm for Dynamic Routing of Capsules

#### Step 1:

To put it another way, if " $v_j$ " and " $p_j$ " refer to capsule  $j$ 's vector weighted sum of all previous layer capsule transfers to this layer, then capsule  $j$  is

$$v_j = \frac{\|p_j\|^2}{1 + \|p_j\|^2} \frac{p_j}{\|p_j\|}$$

#### Step 3:

The prediction vectors are defined as follows;

$$p_j = \sum_i c_{ij} \hat{u}_{j,i}$$

#### Remarks:

To obtain  $p_j$ , one needs to first calculate 'prediction vector'

$$\hat{u}_{j,i} = m_{i,j}u_i$$

" $u_i$ " is the output of the previous layer, where " $m_{i,j}$ " is the weighted matrix between capsules  $I$  in the previous layer and capsule " $j$ " in the current layer. Then,

#### Step 4:

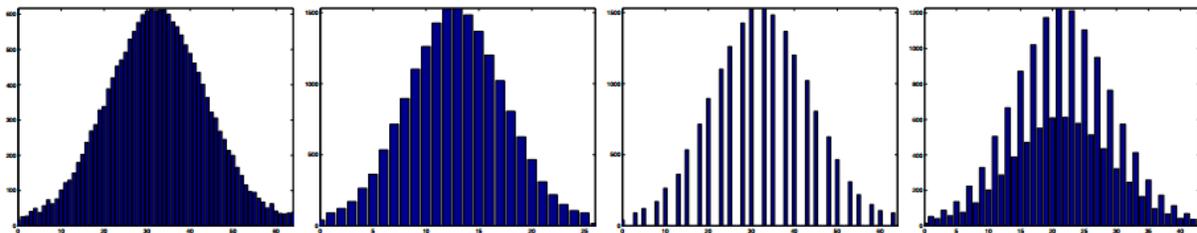
The expected loss can be used as a cross entropy loss function to train in our proposed model with the following equation,

$$L_E = -(y \log(\hat{y}) + (1 - y) \log(1 - (\hat{y})))$$

Here  $L_E$  is the entropy loss function for capsule network model.

## 5. Results & Discussion

In this part, a series of tests demonstrate that the deep learning approach we propose for detecting image fraud works. Furthermore, we compare the proposed method to three prominent image fraud detection algorithms using standard test datasets [32]. The histogram may be generated by the suggested method at an earlier stage in order to determine whether or not the pictures are faked. As a result of the quantization effect being found in the histogram element, we may infer that the histogram element is false.



**Figure 5.** Histogram to Detect Forgery Image Classification at Initial

Nonetheless, if this impact is not detected, it is impossible to draw certain conclusions about the picture since it might, for example, be generated from a RAW file, modified in a graphic editor, and saved as a JPEG file all at the same time without any further processing. Figure 5 shows the histogram mapping for input images. In order to assess the benefit of utilising random noise, we have evaluated our suggested technique both with and without random noise, including the following types:

1. Capsule-Forensics
2. Capsule-Forensics Noise



**Figure 6.** Forged Image Detection by our Proposed Algorithm

The random noise was produced from a normal distribution with a range of values ranging from 0 to 0.1, and it was utilised in both the training and testing phases of the experiment. The dynamic routing method was run through two iterations ( $r = 2$ ) in order to provide the most efficient result. Figure 6 shows results obtained by proposed algorithm. Table 1 shows computed performance metrics.

**Table 1.** computed performance metrics

S.No	Methods	Accuracy	Recall	Precision	HTER (%)
1	MesoNet Network	78.23%	76.14%	71.98%	12.47
2	MesoInception	88.12%	80.73%	84.34%	1.97
3	Capsule Forensic	93.1%	94.21%	92.5%	0.36
4	Capsule-Forensic-Noise	94.9%	97.41%	95.4%	0.001
5	Proposed Intelligent capsule network	95.4%	97.40%	96.4%	0.010



**Figure 6.** Overall Performance Chart

Each data set is divided into six equal-sized groups, with one-sixth of non-repetitive genuine and forged pictures randomly chosen from the data set, to assess the detection performance of the suggested and other methods involved. In both cases, the false acceptance rate shows outliers that have not been identified, while the false rejection rate includes a number of excellent arguments that have been incorrectly categorized as outliers. The performance metrics are following;

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$HTER = \frac{FRR + FAR}{2}$$

The entire performance chart except HTER metrics has shown in Figure 6. It has been shown that the accuracy of proposed intelligent capsule network is superior to that of other currently available techniques.

## 6. Conclusion

The histogram mapping method, which is the foundation of our suggested intelligent approach, allows us to give accurate or false information on input pictures at the time of their creation. Then, using a dynamic capsule network, the input picture will be analyzed to determine if it is a counterfeit or a genuine image. The results of our suggested method for identifying the counterfeit picture are excellent and efficient, as shown in the following table. Furthermore, the accuracy is much greater than that of other conventional techniques. The future study will be focused on researching capsule networks for synthetic forged image detection as well as generic countermeasures for unknown spoofing assaults to improve their detection capabilities. A major emphasis of future research will be on assessing the ability of the suggested approach to withstand adversarial machine assaults, particularly concerning the proposed random noise at test time, and on improving that ability. It will also concentrate on improving the robustness and efficiency of

our proposed methodology for dynamic attacks that must be increasing the synthetic detection issues of the forensic community [33, 34].

## References

- [1] Silva, Jesús, Noel Varela, Fabio E. Mendoza-Palechor, and Omar Bonerge Pineda Lezama. "Deep learning of robust representations for multi-instance and multi-label image classification." In International Conference on Image Processing and Capsule Networks, pp. 169-178. Springer, Cham, 2020.
- [2] Kumar, T. Senthil. "Study of Retail Applications with Virtual and Augmented Reality Technologies." *Journal of Innovative Image Processing (JIIP)* 3, no. 02 (2021): 144-156.
- [3] Basha, S., Dubey, S. R., Pulabaigari, V. & Mukherjee, S. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*. 378, 112–119 (2020).
- [4] Adam, Edriss Eisa Babikir. "Survey on Medical Imaging of Electrical Impedance Tomography (EIT) by Variable Current Pattern Methods." *Journal of ISMAC* 3, no. 02 (2021): 82-95.
- [5] Janeera, D. A., and S. Sasipriya. "A Brain Computer Interface Based Patient Observation and Indoor Locating System with Capsule Network Algorithm." In International Conference on Image Processing and Capsule Networks, pp. 258-268. Springer, Cham, 2020.
- [6] Wei, Q., Jiang, Y. & Chen, J. Machine-learning solver for modified diffusion equations. *Phys. Rev. E* 98, 053304 (2018).
- [7] Kurup, R. Vimal, M. A. Anupama, R. Vinayakumar, V. Sowmya, and K. P. Soman. "Capsule network for plant disease and plant species classification." In International conference on computational vision and bio inspired computing, pp. 413-421. Springer, Cham, 2019.

- [8] Z. Wu, R. K. Das, J. Yang, and H. Li, "Light convolutional neural network with feature genuinization for detection of synthetic speech attacks," in *Inter speech 2020*, 2020.
- [9] Koppa, Anant, Siddharth Kailasam, M. Varun, and Iresh Hiremath. "Pediatric Bone Age Detection Using Capsule Network." In *Inventive Computation and Information Technologies*, pp. 405-420. Springer, Singapore, 2021.
- [10] Y. Yang, H. Wang, H. Dinkel, Z. Chen, S. Wang, Y. Qian, and K. Yu, "The SJTU Robust Anti-Spoofing System for the ASVspoof 2019 Challenge," in *Inter speech 2019*, 2019, pp. 1038–1042.
- [11] Senthilkumar, D., C. Akshayaa, and D. George Washington. "Efficient Deep Learning Approach for Multi-label Semantic Scene Classification." In *International Conference on Image Processing and Capsule Networks*, pp. 397-410. Springer, Cham, 2020.
- [12] D. Cozzolino, G. Poggi and L. Verdoliva, "Splicebuster: A new blind image splicing detector," in *IEEE Workshop on Information Forensics and Security (WIFS)*, 2015, Rome, 2015, pp. 1–6.
- [13] Manoharan, J. Samuel. "Capsule Network Algorithm for Performance Optimization of Text Classification." *Journal of Soft Computing Paradigm (JSCP)* 3, no. 01 (2021): 1-9.
- [14] Yohanandan, S. A., Dyer, A. G., Tao, D. & Song, A. Saliencypreservation in low-resolution grayscale images. *Eur. Conf. Comput. Vis. (ECCV)*. 6, 235–251 (2018).
- [15] Balasubramaniam, Vivekanadam. "Artificial Intelligence Algorithm with SVM Classification using Dermoscopic Images for Melanoma Diagnosis." *Journal of Artificial Intelligence and Capsule Networks* 3, no. 1: 34-42.
- [16] A. Gomez-Alanis, A. M. Peinado, J. A. Gonzalez, and A. M. Gomez, "A light convolutional GRU-RNN deep feature extractor for ASV spoofing detection," in *Proc. Interspeech 2019*, 2019, pp. 1068–1072.

- [17] Adam, Edriss Eisa Babikir, and A. Sathesh. "Construction of Accurate Crack Identification on Concrete Structure using Hybrid Deep Learning Approach." *Journal of Innovative Image Processing (JIIP)* 3, no. 02 (2021): 85-99.
- [18] Y. Q. Shi, C. Chen, and W. Chen, "A natural image model approach to splicing detection," in *Proceedings of the 9th workshop on Multimedia & security. (MM & Sec)*, Dallas, TX, USA, 2007, pp. 51–62.
- [19] Joon Son Chung, Amir Jamaludin, and Andrew Zisserman, "You said that?," arXiv preprint arXiv:1705.02966, 2017.
- [20] Supasorn Suwajanakorn, Steven M Seitz, and Ira Kemelmacher-Shlizerman, "Synthesizing obama: learning lip sync from audio," *ACM TOG*, 2017.
- [21] Wonjun Kim, Sungjoo Suh, and Jae-Joon Han, "Face liveness detection from a single image via diffusion speed model," *IEEE TIP*, 2015.
- [22] Jianwei Yang, Zhen Lei, and Stan Z Li, "Learn convolutional neural network for face anti-spoofing," arXiv preprint arXiv:1408.5601, 2014.
- [23] David Menotti, Giovani Chiachia, Allan Pinto, William Robson Schwartz, Helio Pedrini, Alexandre Xavier Falcao, and Anderson Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE TIFS*, 2015.
- [24] Aziz Alotaibi and Ausif Mahmood, "Deep face liveness detection based on nonlinear diffusion using convolution neural network," *Signal, Image and Video Processing*, 2017.
- [25] X. Zhao, S. Wang, S. Li and J. Li, "Passive Image-Splicing Detection by a 2-D Noncausal Markov Model," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 2, pp. 185–199, Feb. 2015.
- [26] S. Lyu, X. Pan, and X. Zhang, "Exposing Region Splicing Forgeries with Blind Local Noise Estimation," *International Journal of Computer Vision*, vol. 110, no. 2, pp. 202–221, 2014.

- [27] G. Muhammad, M. Al–Hammadi, M. Hussain, G. Bebis, “Image forgery detection using steerable pyramid transform and local binary pattern,” *Machine Vision and Applications*, pp. 1–11, 2013.
- [28] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” in *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017, pp. 3856–3866.
- [29] Aziz Alotaibi and Ausif Mahmood, “Deep face liveness detection based on nonlinear diffusion using convolution neural network,” *Signal, Image and Video Processing*, 2017.
- [30] Dhaya, R. "Hybrid Machine Learning Approach to Detect the Changes in SAR Images for Salvation of Spectral Constriction Problem." *Journal of Innovative Image Processing (JIIP)* 3, no. 02 (2021): 118-130.
- [31] Raj, Jennifer S. "Security Enhanced Blockchain based Unmanned Aerial Vehicle Health Monitoring System." *Journal of ISMAC* 3, no. 02 (2021): 121-131.
- [32] Chen, Joy Iong Zong, and Joy Iong Zong. "Automatic Vehicle License Plate Detection using K-Means Clustering Algorithm and CNN." *Journal of Electrical Engineering and Automation* 3, no. 1 (2021): 15-23.
- [33] Smys, S., and Wang Haoxiang. "Naïve Bayes and Entropy based Analysis and Classification of Humans and Chat Bots." *Journal of ISMAC* 3, no. 01 (2021): 40-49.
- [34] Ranganathan, G. "A Study to Find Facts Behind Preprocessing on Deep Learning Algorithms." *Journal of Innovative Image Processing (JIIP)* 3, no. 01 (2021): 66-74.

### **Author's biography**

**J. Samuel Manoharan** is a professor in the Department of Electronics and Communication Engineering at Sir Isaac Newton College of Engineering and Technology, India. His area of research includes Digital Image and Signal Processing, Data Security and Cryptography, Embedded Systems, Biomedical Instrumentation, Artificial Intelligence, Robotics, Deep Learning,

Cognitive Science, Ad-hoc Networks, Artificial Neural Network, Evolutionary Computing,  
Speech Recognition and Autonomous Systems.