

## An approach to partial occlusion using deep metric learning

Chethana Hadya Thammaiah<sup>1</sup>, Trisiladevi Chandrakant Nagavi<sup>2</sup>

<sup>1</sup>Department of Computer Science & Engineering, JSS Science & Technology University, Mysuru, India

<sup>1,2</sup>Department of Computer Science & Engineering, Vidyaardhaka College of Engineering, Mysuru, India

---

### Article Info

#### Article history:

Received May 12, 2021

Revised Oct 27, 2021

Accepted Nov 10, 2021

---

#### Keywords:

Deep metric learning

Disguised faces in the wild

Euclidean distance

Face recognition

Histogram of gradients

---

### ABSTRACT

The human face can be used as an identification and authentication tool in biometric systems. Face recognition in forensics is a challenging task due to the presence of partial occlusion features like wearing a hat, sunglasses, scarf, and beard. In forensics, criminal identification having partial occlusion features is the most difficult task to perform. In this paper, a combination of the histogram of gradients (HOG) with Euclidean distance is proposed. Deep metric learning is the process of measuring the similarity between the samples using optimal distance metrics for learning tasks. In the proposed system, a deep metric learning technique like HOG is used to generate a 128d real feature vector. Euclidean distance is then applied between the feature vectors and a tolerance threshold is set to decide whether it is a match or mismatch. Experiments are carried out on disguised faces in the wild (DFW) dataset collected from IIT Delhi which consists of 1000 subjects in which 600 subjects were used for testing and the remaining 400 subjects were used for training purposes. The proposed system provides a recognition accuracy of 89.8% and it outperforms compared with other existing methods.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

### Corresponding Author:

Chethana Hadya Thammaiah

Department of Computer Science & Engineering, JSS Science & Technology University

Campus Roads, University of Mysore Campus, Mysuru, Mysuru, Karnataka 570006, India

Email: chethanaht2015@gmail.com

---

## 1. INTRODUCTION

Face is a unique part in each and every individual which performs passive identification in one-to-many environments. Occlusion refers to extraneous objects which hide the face e.g.: face partially covered with hat, sunglasses, scarf and beard [1]. The different types of occlusions occur in face is illustrated in Figure 1. The partial occluded features does not exist more in normal face recognition because the face is captured in a very good scenario with more lightning conditions [2]. But in forensic face recognition, face is captured in a worst scenario with poor lightning conditions.

Presence of these partial occluded features in face makes the process of identification a difficult task which in turn affect the overall performance of the system. Thus, the presence of partial occluded features in face plays a significant role in forensics. If an individual has committed a crime wherein which the face is captured under worst scenario having partial occluded features, then it becomes a very difficult and challenging task to recognize the face of a suspect who involved in criminal activities.

As a result, face recognition having partial occluded features in the field of forensics is a more challenging task than normal face recognition [3]. The main idea behind this work is to recognize partially occluded faces using deep metric learning. Some of the challenges which can be addressed in the field of forensic face recognition are obstructions on face (partial occlusion), facial marks, face captured in surveillance camera captured at the time of crime. Hence in this paper, an attempt is made in this direction to

address the problem of partial occlusion in forensic face recognition. The key contributions of this work are stated:

- An approach based on the combination of histogram of gradients (HOG) with Euclidean distance is proposed
- Proposed system is experimented on disguised faces in the wild (DFW) dataset which contains 1000 subjects with 1, 11,157 images
- Performance of proposed system is compared with other existing methods
- Results obtained from proposed system outperforms with other existing methods in terms of recognition accuracy

The organization of the paper is as follows: Section 1 presents introduction to face recognition and challenges in forensic face recognition. Section 2 discusses literature review on partial occlusion in face recognition. A detailed discussion on proposed system is presented in section 3. Experimental results are discussed in section 4. Finally, section 5 portrays the concluding remarks.



Figure 1. Illustrates the different types of occlusions in face

## 2. LITERATURE SURVEY

In this section, a brief literature review on partial occlusion-based face identification and methods applied for partial occlusion-based face identification is discussed. A method based on multi cascaded convolutional neural network (CNN) is proposed [4] which performs the fusion of features at different levels producing feature vectors. Then classification is performed by applying cosine distance. Deep disguise recognizer network (DDRNET) model is proposed which uses the inception network to train the preprocessed images and similarity metric is calculated which can be used for the classification [5].

Singh *et al.* [4] proposed a deep convolutional neural network (DCNN) based approach for recognition. Here features were trained using two different networks and resulting features were used for recognition. Occluded face recognition consists of both global and local features. Mixing of local features will significantly improve the recognition accuracy in terms of partial occlusion [6]. Azeem *et al.* [6] proposed a lateral subspace strategy to acquire the local features and linear discriminant analysis is used to identify inter and intra class variations using weighted process. The experiment is carried out on augmented reality (AR), Japanese female facial expression (JAFPE), the face recognition technology (FERET) and Extended Cohn-Kanade (CK+) dataset which provides promising results [7].

---

*An approach to partial occlusion using deep metric learning (Chethana Hadya Thammaiah)*

Wen et al. proposed [8] a structure based occlusion coding method which includes structured dictionary and structure sparsity. This structure occlusion coding is used to remove the occlusion from non-occluded images and thereby classification is performed for recognition. Min et al. [9] proposed an approach which is used to detect the occlusions in face and selective local Gabor patterns were applied on non-occluded facial regions. Experimental results proves that the proposed method [10] out performs the state of the art compared to other existing methods. An approach based on weighted matrix is proposed in [11] which consist of two phases: occlusion detection and recognition. In the first phase, partial occlusions are detected using feature-based approach and support vector machine (SVM) with Euclidean nearest neighbor is used for face recognition.

Jozef et al. [12] proposed a large scale deep learning method which uses CNN to detect the faces. Experiment is carried out on annotated faces in the wild (AFW) , face detection dataset, and the benchmark (FDDB) dataset. Experimental results envisage that large scale deep learning method on AFW dataset shows a significant improvement in the overall recognition accuracy.

Devi [13] presents a comparative study of different methods like principal component analysis, K-Principal component analysis and SVM on partial occluded images. Jozef et al. [12] uses Viola jones algorithm for face detection in which these occluded images are sent as an email to the authorized person.

Wu et al. [14] proposed a novel object tracking algorithm called partial occlusion by background alignment (POBA). POBA tracker outperforms the six other trackers on OTB2015 benchmark dataset. An approximation algorithm based on greedy approach is presented in [15]. In the first step, string generation algorithm is designed to convert the face in to string of characters. Then approximation algorithm based on greedy approach is designed to perform string matching. Experiments are carried out on FEI, IAB, ORL and extended Yale-B dataset which shows significant improvement in the recognition accuracy.

Athreya et al. [16] proposed an approach based on the integration of compositional models with DCNN thereby generating a unified deep model which is robust to partial occlusion. A survey on partially occluded faces is presented in [15] which describe the different methods on partially occluded faces. From the comparative study it is observed that REGT provides promising results and performance is better compared with other existing methods. Most of the works [12]–[16] are carried out on partial occlusion based face recognition but none of the method has encountered a combination of HOG with Euclidean distance. To address this problem, an attempt is made in this direction to improve the recognition accuracy.

### 3. PROPOSED SYSTEM

In this section, a brief discussion on the implementation of proposed system for partial occlusion-based face recognition is presented. Proposed system is experimented on DFW dataset [16] which consists of 1000 subjects with 1,11,057 images. The dataset is divided in to testing and training set in which 600 subjects are used for testing and remaining 400 subjects are used for training.

Proposed system uses deep metric learning technique which generates a real-valued feature vector as an output. For the dlib facial recognition network, the output is 128-d feature vector which is used to quantify the face. The network is pre-trained using triplets, two images are faces of one subject and the third image is a random face from the dataset and is not the same subject. The metric quantified from the face is serialized and stored in the database.

If a new face having partial occlusion features needs to be recognized and the corresponding identity has to be verified, then the same HOG model is used to quantify the new face into a 128-d feature vector. This vector is then compared with the stored face encodings using a euclidean distance between the feature vectors [17].

A tolerance threshold is set before to perform the comparison process and a decision is made to check whether it is a match or mismatch. The subject with most matches is considered to be the recognized face in the image sample that is being tested. CNN based model or a 5-point HOG model using dlib is used to perform the quantification process. The HOG model is more successful, faster and more tolerant towards partial facial occlusion and hence it is employed in the proposed system. Proposed system consists of six phases: image acquisition, pre processing, face quantification, feature vector similarity detection, serialization and post processing which is depicted in the Figure 2.

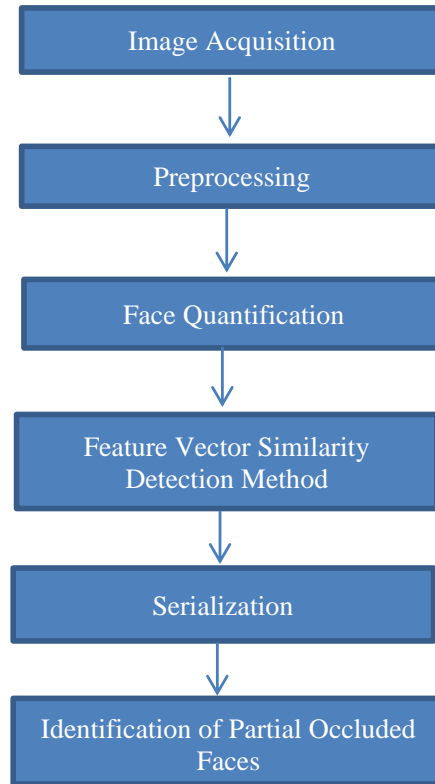


Figure 2. Architecture of the proposed system.

### 3.1. Pre-processing stage

In the pre-processing stage, face detection from standard openCV methods using local binary pattern histogram (LBPH) are employed to isolate the faces and extract them for further steps. Since the faces do not have much sharp contrast features so there is no need to perform extensive sharpening filters or unsharp masking. All the images are converted in to RGB color scheme from a BGR color coding scheme because of compatibility and uniformity reasons. The detailed explanation of image acquisition step is provided in the experimental results section.

### 3.2. Face quantification Method

The HOG model is trained with the images having a dataset of 3 million images using a triplet training method where in two images of the subject and one false positive is used in training the model. This model has an accuracy of 99.38% on labelled faces in wild (LFW) dataset and so transfer learning method is applied to quantify the faces in DFW dataset. The HOG model is a ResNet network which consists of 29 convolutional layers where in few layers are removed and the number of filters per layer is reduced by half. The network is trained with randomly initialized weights and structured metric loss is used which tries to project all the identities into non-overlapping balls of radius 0.6. The loss is basically a type of pair-wise hinge loss that runs over all pairs in a mini-batch and includes hard-negative mining at the mini-batch level. The threshold for similarity index is derived empirically to maximize the verification accuracy keeping in mind the loss projection are non-overlapping balls of radius 0.6.

### 3.3. Feature vector similarity detection method

Even though there are multiple methods exist to generate the similarity of the 128-d feature vectors. In proposed system, euclidean distance is employed to generate 128-d feature vectors. The euclidean distance  $d(p, q)$  in two dimensions is generalized to 128-d as  $d(p, q, n=128)$  where  $p$  and  $q$  are the coordinates,  $n$  represents the number of feature vectors. This method is used to identify the proximity of the sample feature vector and the reference feature vector [14] which is shown in the (1).

$$\sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

### 3.4. Serialization

Pickle module in python is used to serialize and deserialize the feature vectors. This pickle module is the most fundamental and powerful algorithm for serialization and deserialization process. “Pickling” is the process whereby a Python object hierarchy is converted into a byte stream and “unpickling” is the inverse operation, whereby a byte stream is converted back into an object hierarchy. In proposed system, serialization process is performed using pickle module which converts Python object of input facial image in to byte stream.

### 3.5. Identification of partial occluded face recognition

It is the last step in proposed system. Here the byte stream generated from the given facial image is used for the recognition of partial occluded faces.

## 4. RESULTS AND DISCUSSION

In this section, a detailed discussion on experiments carried out on DFW is presented. The dataset is collected from IIIT Delhi [16] consisting of 1000 subjects and 1, 11, 057 images is indicated in the Table 1.

Experiments were carried out on two scenarios. In the first scenario, input image will be a trained partial occluded face and the probability of matching is good. Results of correct matches in partial occlusion-based recognition are shown in the Figure 3.

Table 1. Statistics of testing and training partition in DFW dataset

dataset name	number of subjects		Challenges	Source	Size
	Testing	Training			
Disguished faces in the wild (DFW)	600	400	Normal, validation, disguised and impersonator images	Image analysis and biometrics Lab @IIIT Delhi	Min- 200*350 dimensions to Max- 1000*1800 dimensions



Figure 3. Recognition of correct matches in partial occlusion based recognition



Second scenario: Input image will be not trained partial occluded face, so the proposed system will detect the nearest matching face and corresponding output will be displayed. In this scenario, even though the input facial image is detected but it is not showing the exact name of the individual. Results of the incorrect matches in partial occlusion-based recognition are shown in the Figure 4.

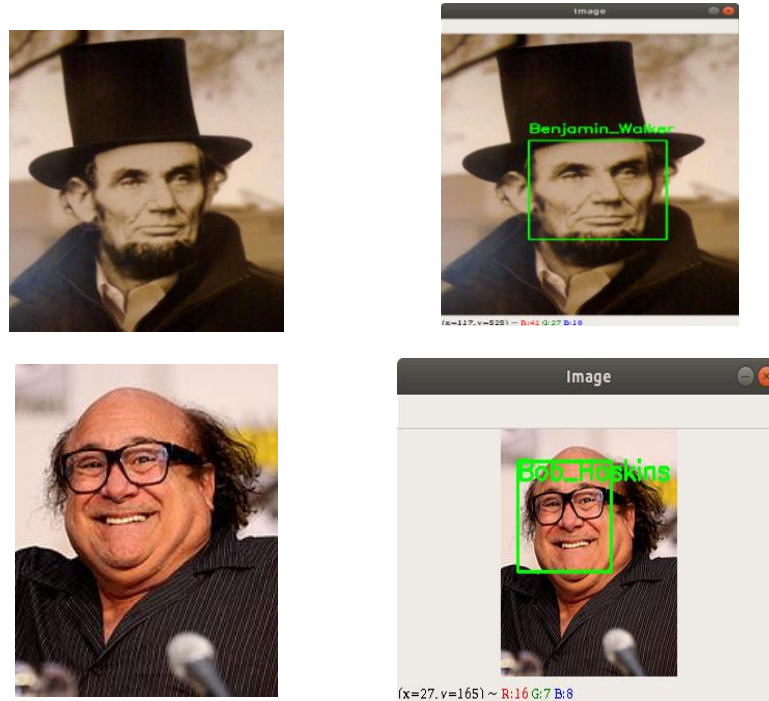


Figure 4. Recognition of incorrect matches in partial occlusion based recognition

Accuracy is one of the metric used to define the closeness of measurements to a specific value. It is also defines as the ratio of total number of samples with correct matches divided by the total number of samples in the dataset. Out of 1000 samples , 898 samples were correctly matched and remaining 102 samples were incorrectly matched . Accuracy is calculated using (2).

$$Accuracy = \frac{Total\ number\ of\ samples\ with\ correct\ matches}{Total\ number\ of\ samples} \tag{2}$$

Proposed system provides a accuracy of 89.8% on DFW dataset and its performance is compared with the other existing algorithms which is illustrated in Table 2.

Table 2. Comparative Study of different methods on partial occlusion based recognition

Sl. No	Authors	Methodology	Recognition accuracy
1	Ashraf Y. A. Maghari et al. [18] (2020)	Region independent component analysis (Reg- ICA Combined)	85%(Sunglass)
2	M Thu et al. [19](2020)	Pyramidal part-based model	81% (Prince of Songkla University dataset)
3	Jaeyoon Jang et al.[20]	Mobile face recognizer	+mask : 88.91% on NIST Mask 0 and 90.89% on NIST Mask 1
4	Proposed system	HOG with Euclidean distance	89.8%

The graphical representation of the comparative study is depicted in Figure 5. In this plot, a value in X-axis denotes the existing methods and proposed system. and a value in Y-axis denotes the recognition accuracy. From the comparative study it is observed that proposed system gives a appreciable recognition accuracy of 89.8% and it outperforms when compared with other existing methods. Proposed system with a combination of HOG with euclidean distance is faster and more tolerant towards partial occlusion based face recognition.

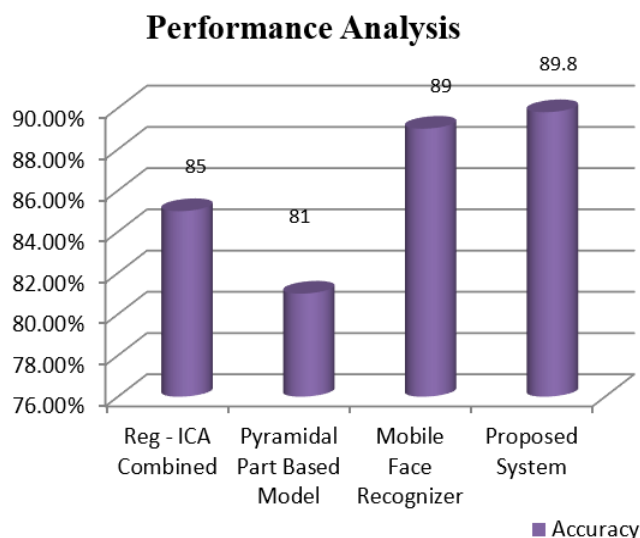


Figure 5. Performance Analysis of existing methods with the proposed system

## 5. CONCLUSION

The main aim of research work is to propose an efficient system for detecting partially occluded images. Partial occlusion face identification is one of the challenges in forensics. An attempt was made in this direction to address the problem of Partial occlusion face identification using a combination of HOG with Euclidean distance. The proposed system works effectively and achieved promising results of 89.8% compared with other existing methods. The proposed system helps in identifying the suspects who involved in criminal activities. Optimizing the threshold for the comparison of faces and more intelligent methods can be applied for calculation of proximity which provides an excellent opportunity for the researchers to carry out the research in the field of partial occlusion-based recognition.

## REFERENCES

- [1] T. B. Patel and J. T. Patel, "Occlusion detection and recognizing human face using neural network," in *Proceedings of 2017 International Conference on Intelligent Computing and Control, I2C2 2017*, Jun. 2018, vol. 2018, pp. 1–4, doi: 10.1109/I2C2.2017.8321959.
- [2] H. T. Chethana and T. C. Nagavi, "A new framework for matching forensic composite sketches with digital images," *International Journal of Digital Crime and Forensics*, vol. 13, no. 5, pp. 1–19, Sep. 2021, doi: 10.4018/IJDCF.20210901.oa1.
- [3] H. T. Chethana and T. C. Nagavi, "A Heterogeneous Face Recognition Approach for Matching Composite Sketch with Age Variation Digital Images," in *2021 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2021*, Mar. 2021, pp. 335–339, doi: 10.1109/WiSPNET51692.2021.9419436.
- [4] M. Singh, R. Singh, M. Vatsa, N. K. Ratha, and R. Chellappa, "Recognizing Disguised Faces in the Wild," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 2, pp. 97–108, Apr. 2019, doi: 10.1109/tbiom.2019.2903860.
- [5] A. Bansal, R. Ranjan, C. D. Castillo, and R. Chellappa, "Deep features for recognizing disguised faces in the wild," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2018, vol. 2018-June, pp. 10–16, doi: 10.1109/CVPRW.2018.00009.
- [6] A. Azeem, M. Sharif, M. Raza, and M. Murtaza, "A survey: Face recognition techniques under partial occlusion," *International Arab Journal of Information Technology*, vol. 11, no. 1, pp. 1–10, 2014.
- [7] F. K. Zaman, A. A. Shafie, and Y. M. Mustafah, "Robust face recognition against expressions and partial occlusions," *International Journal of Automation and Computing*, vol. 13, no. 4, pp. 319–337, Aug. 2016, doi: 10.1007/s11633-016-0974-6.
- [8] Y. Wen, W. Liu, M. Yang, Y. Fu, Y. Xiang, and R. Hu, "Structured occlusion coding for robust face recognition," *Neurocomputing*, vol. 178, pp. 11–24, Feb. 2016, doi: 10.1016/j.neucom.2015.05.132.
- [9] R. Min, A. Hadid, and J. L. Dugelay, "Efficient detection of occlusion prior to robust face recognition," *The Scientific World Journal*, vol. 2014, pp. 1–10, 2014, doi: 10.1155/2014/519158.
- [10] G. N. Priya and R. S. D. Wahida Banu, "Occlusion invariant face recognition using mean based weight matrix and support vector machine," *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 39, no. 2, pp. 303–315, Apr. 2014, doi: 10.1007/s12046-013-0216-3.

- [11] T. Alafif, Z. Hailat, M. Aslan, and X. Chen, "On detecting partially occluded faces with pose variations," in *Proceedings - 14th International Symposium on Pervasive Systems, Algorithms and Networks, I-SPAN 2017, 11th International Conference on Frontier of Computer Science and Technology, FCST 2017 and 3rd International Symposium of Creative Computing, ISCC 2017*, Jun. 2017, vol. 2017-Novem, pp. 28–37, doi: 10.1109/ISPAN-FCST-ISCC.2017.16.
- [12] B. Jozef, F. Matej, O. L'uboš, O. Miloš, and P. Jarmila, "Face recognition under partial occlusion and noise," in *IEEE EuroCon 2013*, Jul. 2013, pp. 2072–2079, doi: 10.1109/EUROCON.2013.6625266.
- [13] A. T. Devi, "A REVIEW ON FACE DETECTION UNDER OCCLUSION BY FACIAL ACCESSORIES," *International Research Journal of Engineering and Technology*, vol. 4, no. 12, pp. 672–674, 2017.
- [14] F. Wu, C. M. Vong, and Q. Liu, "Tracking objects with partial occlusion by background alignment," *Neurocomputing*, vol. 402, pp. 1–13, Aug. 2020, doi: 10.1016/j.neucom.2020.03.026.
- [15] A. Kortylewski, J. He, Q. Liu, and A. Yuille, "Compositional convolutional neural networks: A deep architecture with innate robustness to partial occlusion," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2020, pp. 8937–8946, doi: 10.1109/CVPR42600.2020.00896.
- [16] S. M. Athreya, S. P. Shreevari, B. S. Aradhya Siddesh, S. Kiran, and H. T. Chethana, "A survey on partially occluded faces," in *Evolutionary Computing and Mobile Sustainable Networks*, vol. 53, 2021, pp. 61–67.
- [17] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa, "Disguised faces in the wild," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2018, vol. 2018-June, pp. 1–9, doi: 10.1109/CVPRW.2018.00008.
- [18] M. Thu and N. Suvonvorn, "Pyramidal Part-Based Model for Partial Occlusion Handling in Pedestrian Classification," *Advances in Multimedia*, vol. 2020, pp. 1–15, Feb. 2020, doi: 10.1155/2020/6153580.
- [19] A. Y. A. Maghari, "Recognition of partially occluded faces using regularized ICA," *Inverse Problems in Science and Engineering*, vol. 29, no. 8, pp. 1158–1177, Aug. 2021, doi: 10.1080/17415977.2020.1845329.
- [20] J. Jang, H. Yoon, and J. Kim, "Improvement of identity recognition with occlusion detection-based feature selection," *Electronics (Switzerland)*, vol. 10, no. 2, pp. 1–12, Jan. 2021, doi: 10.3390/electronics10020167.

## BIOGRAPHIES OF AUTHORS



Chethana H T is working as Assistant Professor in the Department of CS & E Engineering, Vidyavardhaka College of Engineering, and Mysuru. She is currently a PhD scholar in the Department of Computer Science and Engineering at the JSS Science & Technological University, Mysuru. She has received her B E degree in Computer Science and Engineering from Coorg Institute of Technology, Ponnampet during 2011 and M. Tech degree in Software Engineering from Visvesvaraya Technological University during 2015. She has total 5 years of teaching experience. Her research interests include Image Processing, Machine Learning, Deep Learning and Digital Forensics. She has published more than 10 research papers in various international journals and conferences. She is a life member of Indian Society for Technical Education



Dr. Trisiladevi C. Nagavi is working as Assistant Professor in the Department of CS & E, S. J. College of Engineering, JSS Science & Technology University Mysuru was conferred with a Ph.D. in Computer Engineering by VTU Belagavi for her research work on Retrieval of Music through Melody in the year 2017. She completed her M. Tech in Software Engineering and secured "Second Rank" from VTU Belagavi during 2004. She obtained her B.E. in CS&E from Karnataka University Dharwad in the year 2000. She has an expertise in the area of Audio, Music, Speech and Image Signal Processing, Information Retrieval, Machine Learning and Digital Forensics. Her research outcome resulted in an AndroidApp to play favourite tunes. She is an Active member of IEEE India Special Interest Group on Communications Disability and worked on Real Time Text to Speech Conversion and Translation System in Collaboration with All India Institute of Speech and Hearing (AIISH) Mysore and IEEE Standards Association. Reviewer for peer reviewed Journals and conferences