

Ensemble Machine Learning for Marker-Free GAIT-Based Human Authentication

Amogha B^{1*}, Rohini Deshpande², Prameela Kumari N²

¹ Reva University, Rukmini Knowledge park, Yelahanka, Bengaluru- 560064, Karnataka, India
r20pec04@ece.reva.edu.in - ORCID: 0000-0001-9782-5655

² Reva University, Rukmini Knowledge park, Yelahanka, Bengaluru- 560064, Karnataka, India
rohinideshpande@reva.edu.in ORCID: 0000-0002-2419-5223

Article Info

Article history:

Received Jul 30, 2025

Revised Feb 26, 2026

Accepted Mar 18, 2026

Keywords:

I-Gait

Authentication

Mediapipe

Machine learning

Morphology

ABSTRACT

The pattern of walking is called as GAIT. This pattern can be used to authenticate persons from a distance. I-GAIT is a method where Indian people are identified and authenticated using a walking pattern. Biometric-based authentication systems are designed to authenticate people using their behavioural and physical patterns. After the COVID-19 pandemic, people are adopting to contactless authentication systems. GAIT is in great demand as authentication is contactless. This paper proposes a methodology to authenticate 240 persons who walked indoors and outdoors in a complex environment. Each person is made to walk in three different situations, such as normal walking (NW), holding a bag (HG), and wearing a coat (WC). Authentication is achieved by calculating the distance between two, four, five and cross-body human joints and morphological features. The distance of the human body is determined by extracting the landmarks from the colour image of a person using mediapipe. Features are trained using five machine learning methods such as random forest, XGboost, LightGBM, gradient boosting and logistic regression. Human recognition is performed by using hard voting, where the majority voted human class is provided as the final predicted class. Authentication accuracy to indoor database is 83%, and for the outdoor database is 86%.

Copyright © 2026 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Amogha B,

Department of Electrical and Computer Engineering,

Reva University,

Rukmini Knowledge park, Yelahanka, Bengaluru- 560064, Karnataka, India

Email: r20pec04@ece.reva.edu.in

1. INTRODUCTION

Biometric-based human GAIT prediction and recognition have gained importance in the past few years as they can be introduced in various places. These systems can be placed in airports, railway stations, bus stands, hospitals, border areas and in other public places where authentication is very important and essential [1]. Reasons for accepting GAIT are as follows,

- Authentication is possible from a distance
- Authentication is performed for low-resolution images
- The authentication process is contactless
- CCTV videos are enough to authenticate and identify persons.

GAIT authentication is widely employed due to the above-stated reasons. It is useful during pandemic times as authentication is from a distance and contactless. When the number of input devices is fewer, observable people act normally and do not hide their biological characteristics. Compared to other biometrics methods such as face, fingerprint, thumb, voice, iris [2], etc. GAIT has numerous advantages, such as authentication is possible at a lower distance and detection from a longer distance is possible. In addition to recognition, same method can be extended in detecting the prevention of crime, forensic, and criminal identification [3]. A limitation of GAIT detection is that it gives false predictions when a person is walking by

wearing a bag, different cloth or wearing bags. To address this issue, many research scholars had proposed two methods first marker-based and marker-free [4][5]. This research is based on the marker-free. It means no accelerometer or sensors are placed on the body, but only provides a decision based on the video of a walking person. Features are extracted from every frame of the video. Mediapipe is employed to identify the landmarks of the body. Features calculated from frames are as follows

- 1) The distance between joints of the body is calculated
 - 2) Morphological and shape features such as area, perimeter, eccentricity, major and minor axis length, solidity, and volume are extracted. Section 3 provides the feature extraction process in detail.
- These two feature sets are combined to obtain 35 features. These features are fed to the four machine-learning approaches to recognise humans.

A general framework for performing a GAIT-based human recognition system is shown in Fig.1. It consists of a collection of datasets, pre-processing images for resizing, segmenting the foreground and background pixels, extracting dominant features, and classifying to specific classes using the classifiers.

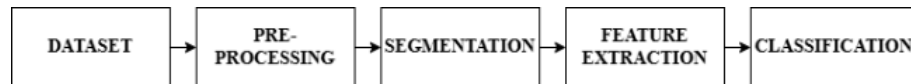


Figure 1. General framework for recognition

Contributions from the paper are as follows,

- Removed the complex objects, such as concrete overhead tankers, banners, vehicles, buildings, and signboards from the frame and captured only the human structure.
- Calculated the distances between the one, second, fourth, fifth and crossover joints of the human body
- Combine the calculated distances and shapes with morphological features. Combined features are fed to four machine-learning algorithms for evaluation
- A hard voting method is used, which decides five machine learning models. The decision is provided based on the majority voting to specific classes from all machine learning methods
- Tuned the models by removing the incorrect frame (redundant frame) from the database to increase the performance metrics

The paper is organized as follows. Section 2 provides an overview of the work carried out by previous research scholars based on GAIT recognition. The next section explains the methodology of the proposed system. Section IV discusses experimental results, and the last section is a conclusion with a future path.

2. Literature Review

Table 1 below formulates the work carried out by the previous researchers on detecting the Human using their walking style. Detection is based on the marker-free and marker-based GAIT method, where the sensors are not placed on the human body. The table consists of the employed dataset, methodology, results achieved, and the gap in the study.

Table 1. Previous research carried on GAIT detection

Year	Dataset	Methodology	Results	Research Gap
2021 [6]	CASIA B	Augmentation is achieved by using three flip operations Feature extraction is performed using NASNET and Inception ResNet V2	89%	Traditional methods are unsuitable for large datasets. Identification from differing views remains a significant difficulty.
2016 [7]	746 adults with an age range of 63 – 89.	Inertial measuring units are placed on the human body XGBoost algorithm is employed. Old age falling is detected Feature: Stride time, length, stance phase, and asymmetry of GAIT	High risk of fall group and poorer cognition (67 to 70% fall risk)	Limited exploration of additional GAIT features for fall prediction.
2016 [8]	CMU MoCap database	Modified linear discriminant analysis with maximum margin criterion	96.2	Need large space for computation for features
2022 [9]	Real-time 50 students. CASIA – B	Real-time video capture, deep learning feature extraction, kurtosis-based selection Correlation-based feature fusion, OaA-SVM-based classification for GAIT recognition	OaA-SVM achieved 95.75% accuracy. Medium KNN achieved 95% accuracy.	Previous studies lacked focus on real-time GAIT recognition. Need for improved fusion approach and larger database.

Year	Dataset	Methodology	Results	Research Gap
2020 [10]	IDNet - 175 and ZJU-GAIT accel - 50 dataset	Supervised and unsupervised approaches Feature extractors and autoencoders	AUC- 95%	--
2018 [11]	ZJU-GAITacc - 153 persons	Features extracted from accelerometer data are made. Features are Euclidian and Manhattan distance,	80%	Feature reduction can increase accuracy
2013 [12]	Casia B database with 124 subjects	Wavelet transforms for feature extraction and analysis. Classification using Nearest Neighbour classification with threshold for identification.	97.98%	No mention of real-time application feasibility.
2020 [13]	OUISIR GAIT D, CASIA B, CMU motion of the body	Convolutional features and histograms of temporal gradients Euclidian distance is considered for the prediction	OUISIR - 99% CASIA B - 97 CMU - 92	Classification is based on Euclidian distance between histogram maximum intensity.
2018 [14]	MOTICON dataset.	An optical motion camera is used for measuring distances. Ground Reaction Forces measured from instrumented foot insoles.	94%	Dependency on the motion camera accuracy.
2015 [15]	Musk molecular dataset	Hanavan model for feature extraction ANN, SVM, and KNN as classifiers	ANN - 84.78% KNN - 84.82% SVM - 92%	--

3. Methodology

The proposed model is able to extract two kinds of features first feature is shape features, such as area, perimeter, eccentricity, major minor axis length, solidity, and volume. Secondly, the calculation of distances of various body joints. These two feature fusion methods are going to enhance the accuracy of the detection rate and the accuracy of the model. Human body anatomy is shown in Fig.2, which helps in joints calculation. Fig.3. The steps involved in the proposed model are as follows,

Step 1: Dataset: Real-time database (I-GAIT) collected in India by considering 240 subjects with four view distances. This dataset provides the segmented human region from the background. Dataset has four viewangles. For every person features from every viewpoint is calculated which increases robustness in detecting persons accurately.

Step 2: Preprocessing: resize the images to 224x224

Step 3: Segmentation: using background subtraction complex objects are separated from the human area in the frame. Background subtraction is a technique that is useful in preprocessing to get rid of the unnecessary background info and to focus on the background of the walking person. The first thing to do is break the video down to single frames. Then, make a background model of the frames that includes the static objects in the scene like the walls and the floor along with some of the objects that stay in the same position. For each frame coming in, can compare it to the model made of the background and the preceding frames, in order to see what is different from the background. The pixels that change from the background move and get classified as background. The model also includes things like the person walking, and that is the generated mask. The original generated mask from the background model might include static objects contains some noise and holes because of the lighting from the background and motion in the video. To get rid of the gaps in the noise from the background, we can use a technique to get rid of the gaps, and fill in the gaps to finish it to get rid of the noise. Then you have a mask with the person and the background model, and the model for the background is also useful in the refining. Pose landmark detection and gait feature extraction can then focus on just the human subject while eliminating the background noise and variations, using this silhouette for additional processing. A basic methodology here is to subtract the static objects in present and previous scene to obtain person silhouette.

Step 4: Extract two kinds of features, the first is shape features, and the second is the distance between joints of the body. Different joints considered are single, two, four, five joints and cross-body joints.

- Single joint (One joint) – Elbow, knee, ankle, shoulder, hip, wrist, neck, jaw
- Two joints – Shoulder to elbow, elbow to wrist, hip to knee, knee to ankle, neck to shoulder
- Four joints – shoulder elbow wrist hand, hip, knee ankle foot, neck shoulder elbow wrist,
- Five joints – Shoulder elbow wrist hand fingers, hip knee ankle foot toes, neck shoulder elbow wrist hand.
- Cross body – left side parts of body (elbow, knee, wrist and ankle) to right side of body.

Reason for choosing only these joints is based on anatomy of human body, single joint means only one bone holding any organ or involved in movement of organ. Need of selecting specific joints are as follows, two joint which represents length of limb, normalization is easy which forms core anthropometric features. Four

joints give one limb chain information and it is going to represent limb chain it extracts trajectory of limb, arm and leg swing, chain motion. Five joint provides body posture information such as inclination, balance in walking and alignment of torso leg. Three joint of body are not necessary for current research as these joints does not provide distance information it gives angle information. Six joints create high dimensional feature and forms complex polygons and also includes redundant information. Hence above selected joints covered all important points of human body to detect person. To address different height of person during authentication, all features relating to distance between joints were normalized to account for differences in body height and camera distance. With regard to joints, the distance calculated was a pairwise one and was normalized using a body distance reference. For this study, length from neck to hip center, served as the normalization factor. The resultant features, post normalization, are scale-invariant, and improved the impact of the camera perspective and subject height variations.

Annotation points on the human body, which help in finding limb and knee lengths are shown in Fig. 2.

Step 5: Calculate the distances between the joints (one to five segments) listed in step 4 and store them in a CSV file. Features stored in CSV files are used for training and testing the four machine-learning models. The decision from all the machine learning is collected. The final prediction is through the hard voting method, i.e., output is the maximum voting for the specific class from all four machine learning algorithms.

Step 6: After training and testing of the model, an accuracy achieved, 55% of wrongly predicted frames are removed from the database. After removing wrongly predicted frames through hard voting achieved an accuracy of 86%. Frames before and after removing are 11764 and 9438, respectively. Thirty-five features are collected from every frame in the video.

Fig.3. illustrates steps employed in the proposed model. Details of the blocks specified in the figure are explained in the next subsections (3.1 to 3.5).

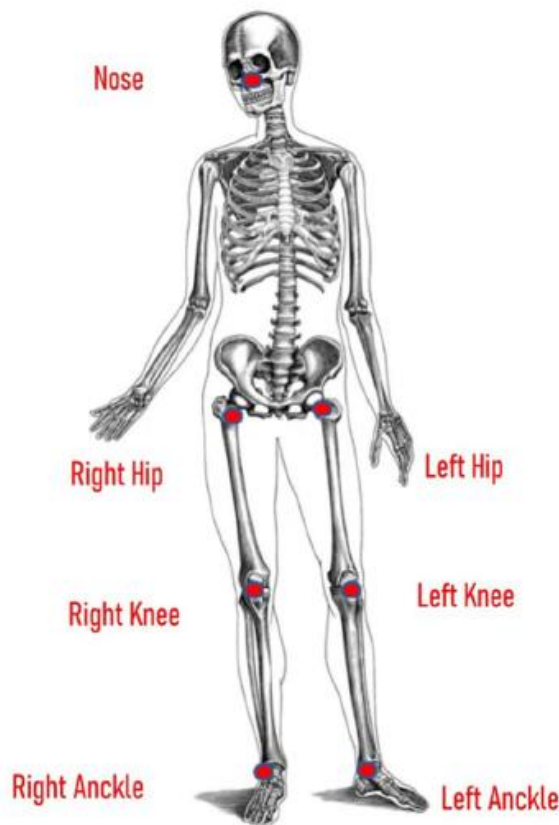


Figure 2. Annotation points for body parts identification [21].

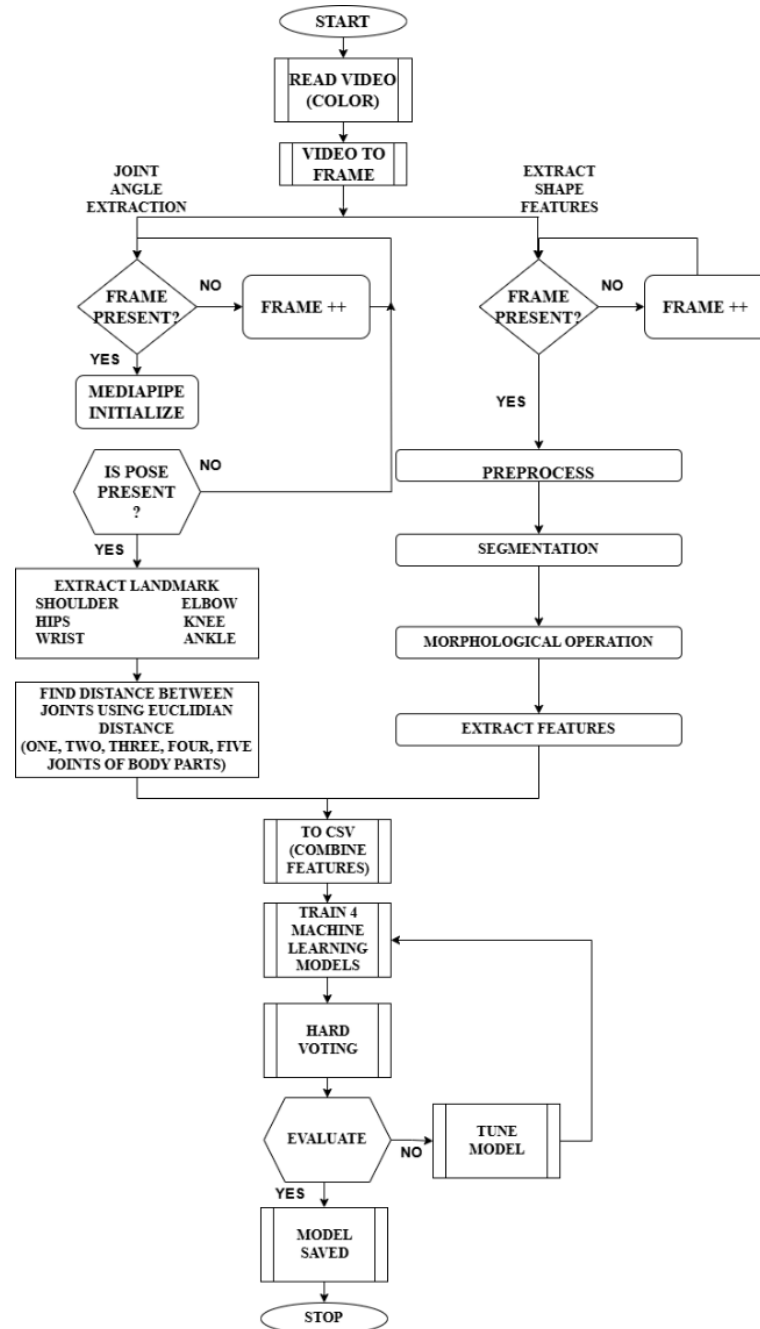


Figure 3. Steps involved in proposed model

3.1 Preparation of Database

The database was prepared by capturing 240 people with four view distances (45, 90, 135, and 180 degrees). The subjects are allowed to walk in indoor and outdoor environments. Videos are captured in natural light where complex objects are present in the background. Every subject walked in three different situations, such as normal walking, walking with a bag, and wearing a coat. Captured videos are stored in individual folders, where every folder has a person walking in four view distances. Fig. 4 shows four view angles (45, 90, 135 & 180) of every subject, and three styles (Normal walking, wearing a coat, and holding a bag) are captured of every person. Videos are collections of frames; the next step is to work on the frames (a collection of images). Hence videos are converted to images. The mathematical model behind the video-to-frame conversion is given below Eq.1.

$$\text{Total number of frames} = \text{Frames per second (30 per second)} * \text{total seconds in video} \quad (1)$$



Figure 4. Three instances walking pattern of every person

3.2 Pre-processing and segmentation

The proposed model has two types of feature collection: one is on distance determination between joints, and the second is the morphological shape feature. Distance detection has not employed a preprocessing step, as it uses the landmarks that are detected directly in colour images. To extract shape features using the morphological features preprocessing step is employed to remove complex background objects. Dataset says that complex objects such as two-wheelers, four-wheelers, over-head tanks, etc.

To remove this background and to retrieve the foreground we are using a mixture of Gaussian distributions [16]. This is the background subtraction algorithm where the moving object is separated from the stationary background. Analogy behind the algorithm is that a pixel in every frame of the video is described as belonging to multiple Gaussian distributions. The rationale is that the background exists more frequently than the moving objects and that the background pixels will be modeled by one or more of the Gaussians, while the moving objects will not be. Eventually, depending on the changes in lighting or slow movements, the model learns and replicates the motion as shown in Fig.5. The steps involved in the segmentation of foreground from background pixels are as follows,

- Finding the pixel intensity:

A mixture of Gaussian 'K' is found by using Eq.2. It is applied on every pixel of video frames.

$$P(I(x,y) = \sum_{i=1}^K \omega_i(x,y) * N(I(x,y); \mu_i(x,y), \Sigma_i(x,y)) \quad (2)$$

ω_i it is the weight of the specified pixel 'i'

$I(x,y)$ is the image with x, y represents the pixel of row x and column y.

Σ_i & μ_i It is the covariance and mean of specified pixel i respectively.

N – represents the Gaussian probability function.

- Identification of pixels as foreground and background

Each Gaussian indicates what the color or intensity of the pixel can be at any one time instant. The background pixels are modelled by at least one or more Gaussian distributions while foreground pixels (moving objects) are said to be outliers.

A pixel can be classified into the background if the value of that pixel falls into either of the calculated Gaussian distributions. If the pixel is not nearer to any distribution, then it is considered foreground. Based on the intensity, a decision is made, and Eq.3 formulates it.

$$I(x,y) - \mu_i(x,y) \leq \text{Threshold} * \sigma_i(x,y) \quad (3)$$

Where, σ_i is the square of standard deviation or variance

The chosen threshold is 127 as it is the middle of the histogram. This threshold is used to generate the binary image where white is the foreground and black represents the background.

Morphological open operation is applied on frames smooth the edges of humans.

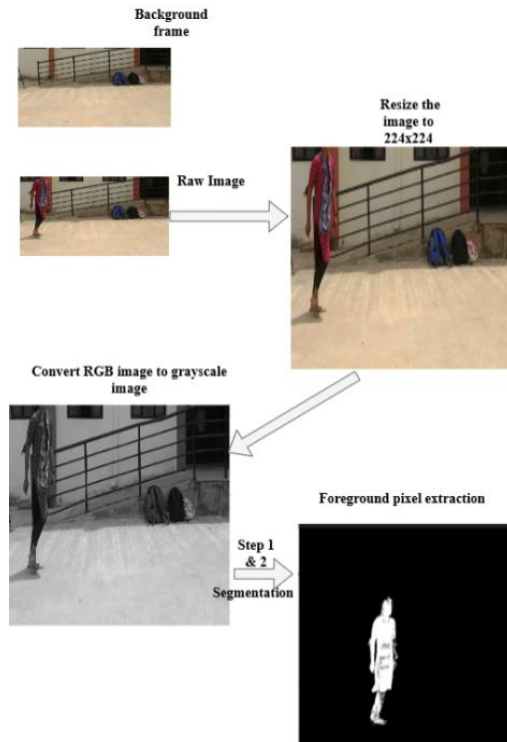


Figure 5. Preprocessing step using the background subtraction method

3.3 Proposed Feature Extraction

We are going to collect two types of features. They are the distance between joints and shape features (volume, solidity, perimeter, area, major and minor axis length).

3.3.1 Distance calculation

An open-source platform called mediapipe is employed to detect the landmarks of the full human body [17]. Landmarks of the body are shown in Fig.6. Once the landmarks are extracted from the frame, we calculate the distances between the joints. Fig.7. shows the joints of the distance calculation. The procedure provides the details of how joints of the body are chosen and the distances between them are calculated.

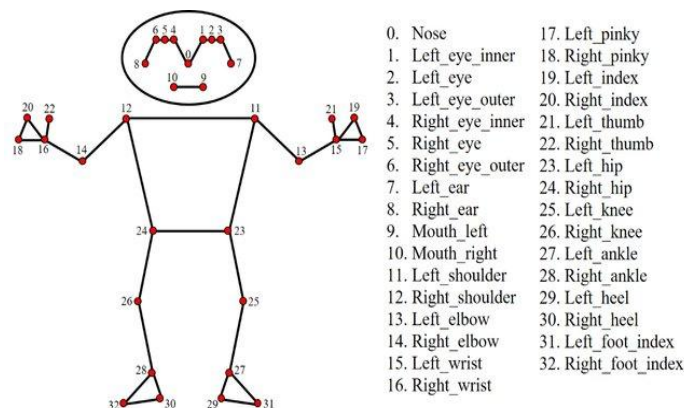


Figure 6. Mediapipe landmarks for full body [18].

The procedure provides the details of how joints of the body are chosen and the distances between them are calculated.

- Selection of specific bones in the body for efficient human recognition: Authentication of humans using GAIT is performed mainly by considering the walking style. To predict the walking style of a person, the major bones considered are the arms, hips, and legs. Hence, we have considered the arm, wrist, hips, ankle, leg, and nasal bone. Different bones that form the segment of bones or joints of bones are provided in Fig.7.
- Selection of major joints of the body (single, two, four, five, and crossover): selection of bone segments and a briefing on how joints of bones are chosen. Also describes the distance calculation between joints.

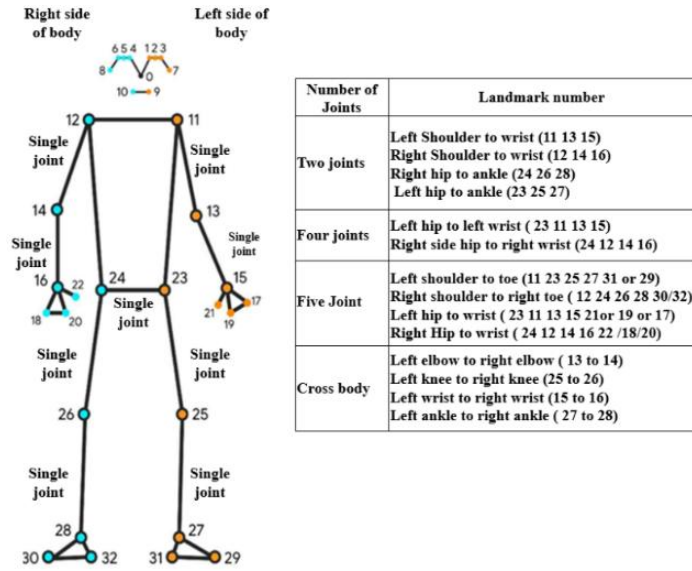


Figure 7. Distance calculation between joints.

Single joint: A joint is thus defined as a single or individual articulation of bones in a body, and forms where two or more bones meet that also facilitate movement. The movement occurring in a single joint means that its interaction is often restricted to some of the freedoms (flexion, extension, rotation, etc.). Single joints are shown in the Fig.10 and explained below.

- Knee Joint: It allows a person to perform extension and flexion, which makes a joint between the femur (thigh) and tibia (shin bone), below figure Fig.8. describes it clearly. In Fig. 7, knee joints are represented by numbers 25 and 26, left and right knees, respectively. The distance formula is applied between the left knee and hip, i.e., 25 and 23. Another pair of distance calculations is between 24 and 26th i.e., between the right knee and hip. The equation for the distance formula is given by Eq. 4.

$$\text{Distance} = \sqrt{\frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2}} \quad [20] \quad (4)$$

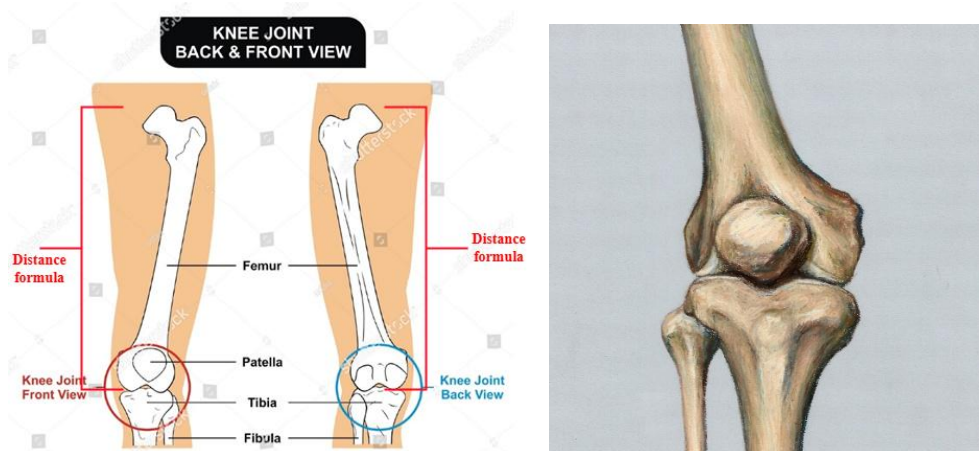


Figure 8. Knee Joint. [19]

Elbow joint: It is the joint that exists between the forearm bone and upper arm bone. Fig. 10 shows the elbow joint of the right hand. Landmarks of elbow joints are four pairs and marked in Fig.9. are 11 to 13, 13 to 15, 12 to 14, and 14 to 16. The distance formula is employed between these pairs of bones to find the elbow joint distance. Fig 11 shows the elbow joints.

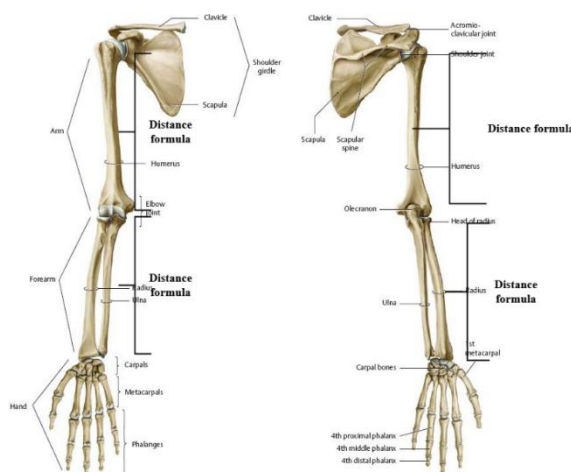


Figure 9. Elbow joints and applied distance formula to joints.[22]

In the same manner, two, four, and five joints of the body are considered to calculate the distance between them. These distances are stored in a CSV file. Cross-body distance is the one where the distance is calculated from the left part of the body to the right. From left to right: knee, elbow, wrist and ankle. A crossbody is going to provide body alignment and symmetry of it. Calculated body joint distance is stored as features, and the label is “person 1 to 240”. Every feature is separated by a comma, and Fig. 10 summarises the workflow of the proposed feature extraction using mediapipe.

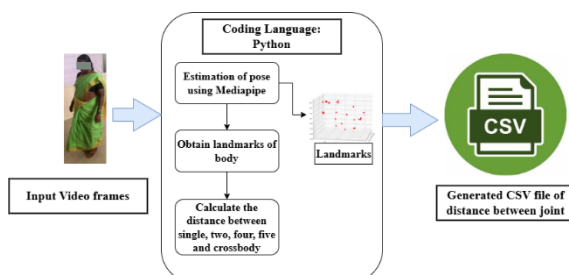


Figure 10. Proposed feature extraction using mediapipe and storage of distance in csv

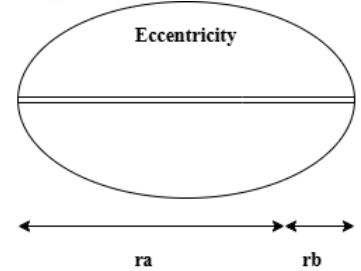
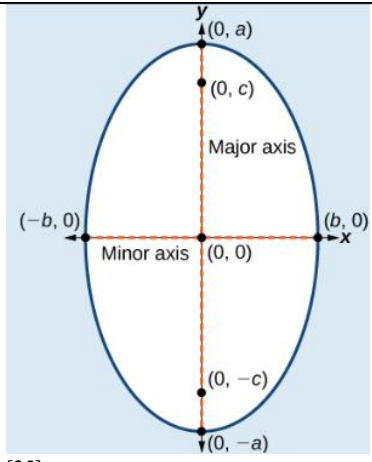
3.3.2 Shape features extraction

In this step, we are applying steps to extract the shape features from the image. Shape features are area, perimeter, eccentricity, major axis length, minor axis length, solidity, and volume. The procedure involved in extracting shape is as follows.

- Preprocessing and segmentation are performed on the grayscale image. Input colour video frames are converted to images and applied to the processing and segmentation steps to obtain a silhouette of the person. The methodology is explained in section 3.2
- Applying morphological operation: Pre-processed and segmented images are used to apply morphological operation. In addition to the subtracted image, another input is the ellipse structural element. This structuring element is called a kernel or matrix. The matrix size employed is 3x3 on the image, and performs the open morphological operation. It removes the foreground pixels that are present on the edge of the image and also removes the smaller noise present in the image [23]. All these operations are performed on binary images.
- Extraction of shape features: A processed binary image is used to extract shape features. Initially, black and white pixels are divided using 8- 8-connected components and labelled. The white pixel is the person's outline, and the black colour is the background. We are interested in white pixels, and this labelled masked image is used to extract the shape features using the library region props of OpenCV. Tab.2. gives a clear

picture of descriptions of features. No feature reduction methods such as PCA or other methods not employed. CSV creation is provided in below section.

Table 2. Shape Features calculation from every frame

Feature	Equation	Details
Area of white pixels	$Area = \sum_{i,j} mask(i,j)$	
Perimeter: Total length of white pixels in only boundary of image	Perimeter = $\sum_i Dist(I_i, I_{i+1})$ Distance between ith and i+1 is found	
Eccentricity [24]	$E = \sqrt{1 - \frac{r_p}{r_a}}$ r_p is length of minor axis r_a is length of major axis	
Major axis and minor axis[24]	$M = a+b$ Where a, b are the distances from 'f' to any point in the ellipse.	
Volume: 3D space which is occupied by the structure or object	Volume = area * mask [0]	[25]

The obtained shape and distance calculated features need to be combined to prepare a label and the features of a specific person. This is achieved by using the csv_writer function, for this function, 28 joint distances and 7 shape features are passed. It is going to merge all features and write 35 features, and a specific label is added to every frame and stored in CSV format.

Prepared CSV files are used for training and evaluating the model. Have employed five machine learning models, such as random forest, XGBoost, LightGBM, gradient boosting, and logistic regression. Once all the models have given their predictions, a hard voting method is employed to give the final prediction. The final prediction is through the hard voting method, i.e., the output is the maximum voting for the specific class from all five machine learning algorithms. Details of machine learning models are as follows,

- Random forest: It is the method that counts the votes from multiple decision trees, which merge all of them to obtain high accuracy and accurate prediction
- XGBoost: XGBoost is a highly effective gradient boosting model that constructs a decision tree-based ensemble for improving the predictive model's accuracy and lessening overfitting. It is fast, scalable, and very appropriate when dealing with different forms of data and dealing with missing values.
- Light GBM: Another gradient boosting framework that is optimised for speed, in particular, for large data sets.
- Gradient boosting: A method for creating models with parameters where the next model tries to refine mistakes made in the preceding model.
- Linear regression: It is a statistical technique used to model a binary outcome, based on a single or several predictor variables, as the probability of that outcome. It outputs values between 0 and 1, where it applies the logistic function to estimate probabilities and classify data points accordingly.
- Hard voting: The aim of using the hard voting method is to avoid the overfitting of the model to the training data. The methodology of hard voting is that the final prediction is selected from the class label that has the majority vote amongst the classifiers.

3.3.3 Procedure for training

The involved procedure for training is explained in this sub-section

- 1) Data Preparation:
The data is preprocessed and then partitioned into a training set and a test set. The models are built with the help of the training data, while the performance is assessed using the testing data.
- 2) Training the Voting Classifier:
All the models in the voting classifier (Random Forest, XGBoost, LightGBM, Gradient Boosting, Logistic Regression) are trained from the same set of training data.
The Voting Classifier then comes up with the overall consensus based on the forecasts given by each Classifier from the majority voting.
- 3) Filtered Data Usage:
The voting_clf_filtered model is a logistic regression model that is trained on the filtered dataset, where all the rows that were misclassified in the voting_clf have been dropped. This filtering is performed in order to improve the overall quality of the model.
- 4) Results and evaluation:
Model results and evaluation are going to be discussed in Section IV. Where the confusion matrix, details on the model, and the employed dataset are provided in detail.

4. Results and Discussion

This section provides details about the results achieved in detecting the person. The robustness of the system is determined by the evaluation metrics. Performance metrics include a confusion matrix, which has rows and columns equal to the number of classes present. It gives information about the accuracy rate where the predicted is plotted against the actual in the x and y axes, respectively. The confusion matrix includes true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The equations below provide equations for calculating the confusion matrix.

TP: It gives the total number of positive class predictions from the model

TN: number of correctly negative class predictions from the model.

FP: It is the count of misclassification of the model, which predicts it as positive, but its original class is negative. It is an incorrect predication of the positive class as negative, for example, if the patient is healthy but the model predicts it as cancer. It is termed a type I error.

FN: It is incorrect to predicate the negative class as positive, for example, a person is cancerous but predicts it as healthy. It is termed a type II error. Type 1 and type 2 errors are dangerous; in type 1, unnecessary treatment is performed, which is a mental and financial burden, but there will be no disease. In type 2, there is a need for treatment, but due to a wrong prediction, the person dies. Hence, the model needs to take FN and FP, which helps in the robustness of the model. The proposed model is evaluated by using the above parameters to calculate accuracy, F1 score, and recall specificity [26]. Tab.3. gives the equations for these performance metrics.

Term	Formula
Accuracy	$\frac{TP + TN}{\text{sum of rows in dataset}} * 100$
Precision	$\frac{TP}{TP + FP} * 100$
Recall	$\frac{TP}{TP + FN} * 100$
F1 score	$\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} * 100$

4.1 Evaluation of the dataset for the outdoor scenario

The dataset has two scenarios, outdoor and indoor, where persons are allowed to walk in these two scenarios. This subsection describes the results obtained for the outdoor scenario. Four view angles were considered (45, 90, 135 & 180 degrees) to capture every person's locomotive patterns. Training time is four hours to complete for all subjects of 120 persons captured in an outdoor scenario. Once the model is trained at the initial stage, it has given 56% of accuracy when 11764 rows were present for training the model. To improve model accuracy, the frame was removed frame which gave the wrong prediction of a specific person. After filtering, the number of rows left over is 9438 which means 2326 frames were removed, which led to misclassification. The criteria for choosing a frame with misclassification is F1score, whose F1 score is less than 70, the frame of a specific person is deleted from the dataset. Hence, after removal, model accuracy was increased to 86%. Thirty-five (35) combined features are extracted for every person, and instance of it is shown in Fig.11. Tab.4. gives the model performance metrics for the outdoor dataset. Fig.12. shows a confusion matrix of the proposed predication.

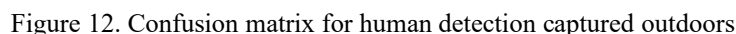


Table 4. Model metrics for Outdoor dataset

Dataset type	Accuracy (%)	Precision (%)	F1 score (%)	Recall (%)
Outdoor (120 persons)	86%	88	81.1	78.5

Evaluation of the model using the indoor dataset

The dataset contains 120 person videos which had been captured in four different view angles (45, 90, 135 & 180 degrees) from four input devices. Using these video models is trained and tested in the ratio of 70:30, respectively. Once the model is trained at the initial stage, it has given 56% of accuracy when 11972 rows were present for training the model. To improve model accuracy, remove the frame that gave the wrong prediction of a specific person. After filtering, the number of rows left over is 9577, which means 2395 frames were removed, which led to misclassification. The criteria for choosing a frame with misclassification is the F1 score, whose F1 score is less than 70; the frame of a specific person is deleted from the dataset. Hence, after removal, model accuracy was increased to 83%. Thirty-five (35) combined features are extracted for every person, and an instance of it is shown in Fig.13. The number of samples before filtering and after filtering is 11972 and 9577, respectively. Fig.14. shows confusion matrix of the proposed model in classifying images into actual versus predicted. Tab.5. gives the model performance metrics for the indoor dataset.

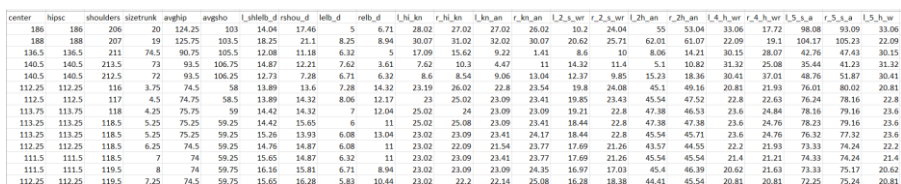


Figure 13. Features extracted for the indoor dataset

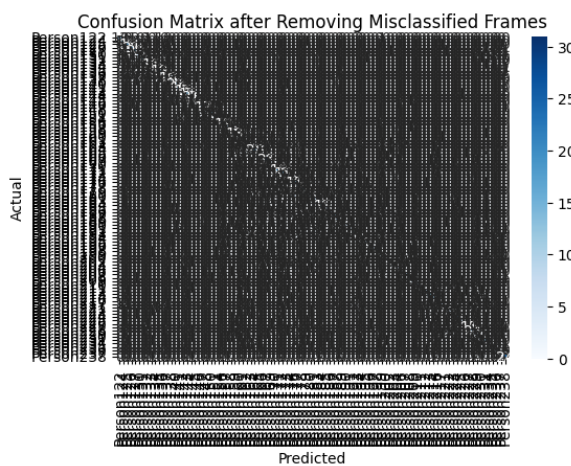


Figure 14. Confusion matrix for the indoor dataset

Table 5. Performance metrics for indoor dataset

Dataset type	Accuracy (%)	Precision (%)	F1 score (%)	Recall (%)
Indoor (120 persons)	83	83.47	82.65%	82.64

5. Conclusion

This paper proposes a methodology to improve the recognition rate in video by detecting the person using a GAIT posture. GAIT means a walking pattern of humans. I-GAIT means authentication of the Indian people using a walking pattern. An Indian-populated dataset is employed to train and test the model. The dataset is formulated for three situations where persons walked by wearing coats, holding a bag, and walking normally. The proposed feature extraction combines distance calculation between joints and shape features to predict humans using the hard voting method. The hard voting method is the ensemble of five machine learning models, such as random forest, XGBoost, LightGBM, gradient boosting, and logistic regression. The flow of the model is that color video frames are converted to binary to extract morphological shape features (area, perimeter, eccentricity, major, minor axis, solidity, volume) and for distance calculation RGB frames are fed to the mediapipe to extract human landmarks, using these landmarks distance between single, two, four, five and crossbody is calculated using the distance formula. Decisions from every machine learning is collectively

fed to a hard voting method. This voting method takes a decision from every machine learning model and gives results that get the maximum vote. Achieved an accuracy of 83% and 86% for indoor and outdoor, respectively in authenticating humans correctly. The time taken for training and testing the model is four hours. The model is executed in Kaggle with a GPU P100.

REFERENCE

- [1] Rani, V., & Kumar, M. (2023). Human GAIT recognition: A systematic review. *Multimedia Tools and Applications*, 82(24), 37003-37037. DOI: <https://doi.org/10.1007/s11042-023-15079-5>
- [2] Khosravy, M., Nakamura, K., Hirose, Y., Nitta, N., & Babaguchi, N. (2022). Model inversion attack by integration of deep generative models: Privacy-sensitive face generation from a face recognition system. *IEEE Transactions on Information Forensics and Security*, 17, 357-372. DOI: <https://doi.org/10.1109/TIFS.2022.3140687>
- [3] Li, H., Qiu, Y., Zhao, H., Zhan, J., Chen, R., Wei, T., & Huang, Z. (2022). GAITSlice: A GAIT recognition model based on spatio-temporal slice features. *Pattern Recognition*, 124, 108453. DOI: <https://doi.org/10.1016/j.patcog.2021.108453>
- [4] Liao, R., Yu, S., An, W., & Huang, Y. (2020). A model-based GAIT recognition method with body pose and human prior knowledge. *Pattern Recognition*, 98, 107069. DOI: <https://doi.org/10.1016/j.patcog.2019.107069>
- [5] Gupta, S. K., & Chattopadhyay, P. (2021). GAIT recognition in the presence of co-variate conditions. *Neurocomputing*, 454, 76-87. DOI: <https://doi.org/10.1016/j.neucom.2021.04.113>
- [6] Faizan, Saleem., Muhammad, Attique, Khan., Majed, Alhaisoni., Usman, Tariq., Ammar, Armghan., Fayadh, Alenezi., Jung-In, Choi., Seifedine, Kadry. (2021). Human GAIT Recognition: A Single Stream Optimal Deep Learning Features Fusion. *Sensors*, DOI: <https://doi.org/10.3390/s21227584>
- [7] Byungjoo, Noh., Changhong, Youm., Eunkyong, Goh., Myeounggon, Lee., Hwayoung, Park., Hyojeong, Jeon., Oh, Yoen, Kim. (2021). XGBoost based machine learning approach to predict the risk of fall in older adults using GAIT outcomes. *Scientific Reports*, DOI: <https://doi.org/10.1038/s41598-021-91797-w>
- [8] Balazia, M., & Sojka, P. (2016, December). Learning robust features for GAIT recognition by maximum margin criterion. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 901-906). IEEE. DOI: <http://dx.doi.org/10.1109/ICPR.2016.7899750>
- [9] Sharif, M. I., Khan, M. A., Alqahtani, A., Nazir, M., Alsubai, S., Binbusayyis, A., & Damaševičius, R. (2022). Deep learning and kurtosis-controlled, entropy-based framework for human GAIT recognition using video sequences. *Electronics*, 11(3), 334. DOI: <https://doi.org/10.3390/electronics11030334>
- [10] Nemes, S., & Antal, M. (2021, May). Feature learning for accelerometer based GAIT recognition. In *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (pp. 479-484). IEEE. DOI: <https://doi.org/10.1109/SACI51354.2021.9465576>
- [11] Maria, De, Marsico., Eduard, Gabriel, Fartade., Alessio, Mecca. (2018). Feature-based Analysis of GAIT Signals for Biometric Recognition - Automatic Extraction and Selection of Features from Accelerometer Signals. 630-637. DOI: <https://doi.org/10.5220/0006719106300637>
- [12] Sabir, A., Al-jawad, N., & Jassim, S. (2013, May). GAIT recognition using spatio-temporal silhouette-based features. In *Mobile Multimedia/Image Processing, Security, and Applications 2013* (Vol. 8755, pp. 194-203). SPIE. DOI: https://ui.adsabs.harvard.edu/link_gateway/2013SPIE.8755E..0RS/doi:10.1117/12.2017950
- [13] Mogan, J. N., Lee, C. P., & Lim, K. M. (2020, March). GAIT recognition using histograms of temporal gradients. In *Journal of Physics: Conference Series* (Vol. 1502, No. 1, p. 012051). IOP Publishing. DOI: 10.1088/1742-6596/1502/1/012051
- [14] S. Sivakumar, A. A. Gopalai, D. Gouwanda and K. H. Lim, "Estimation Of Joint Angle From Ground Reaction Force In Human Gait," 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia, 2018, pp. 623-628, DOI: 10.1109/IECBES.2018.8626650.
- [15] Yang, Chaoyu, Renjie Geng, and Qian Liu. "Gait Feature Recognition Based on Multi-instance Learning." (2015). DOI: <https://pos.sissa.it/259/084/pdf>
- [16] Martins, I., Carvalho, P., Corte-Real, L., & Alba-Castro, J. L. (2018). Bmog: boosted gaussian mixture model with controlled complexity for background subtraction. *Pattern Analysis and Applications*, 21, 641-654. DOI: <https://link.springer.com/article/10.1007/s10044-018-0699-y>
- [17] Kondragunta, J., & Hirtz, G. (2020, July). GAIT parameter estimation of elderly people using 3D human pose estimation in early detection of dementia. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 5798-5801). IEEE. DOI: <https://doi.org/10.1109/embc44109.2020.9175766>
- [18] Chen, K. Y., Shin, J., Hasan, M. A. M., Liaw, J. J., Yuichi, O., & Tomioka, Y. (2022). Fitness movement types and completeness detection using a transfer-learning-based deep neural network. *Sensors*, 22(15), 5700. DOI: <https://doi.org/10.3390/s22155700>
- [19] Lafortune, M. A., Cavanagh, P. R., Sommer Iii, H. J., & Kalenak, A. (1992). Three-dimensional kinematics of the human knee during walking. *Journal of biomechanics*, 25(4), 347-357. DOI: [https://doi.org/10.1016/0021-9290\(92\)90254-x](https://doi.org/10.1016/0021-9290(92)90254-x)
- [20] Lang, S., Murrow, G., Lang, S., & Murrow, G. (1988). The Distance Formula. *Geometry: A High School Course*, 110-122.
- [21] Yousef, R. N., Khalil, A. T., Samra, A. S., & Ata, M. M. (2023). Model-based and model-free deep features fusion for high performed human GAIT recognition. *The Journal of Supercomputing*, 79(12), 12815-12852. DOI: <https://doi.org/10.1007/s11227-023-05156-9>

- [22] “Elbow Image”, available at, <https://www.vroque.co/post/elbow-anatomy-upper-limb-anatomy-anatomy-bones-anatomy-study>
- [23] Comer, M. L., & Delp III, E. J. (1999). Morphological operations for color image processing. *Journal of electronic imaging*, 8(3), 279-289. DOI: <https://doi.org/10.1117/1.482677>
- [24] Tasel, S. F., Mumcuoglu, E. U., Hassanpour, R. Z., & Perkins, G. (2016). A validated active contour method driven by parabolic arc model for detection and segmentation of mitochondria. *Journal of structural biology*, 194(3), 253-271. DOI: <https://doi.org/10.1016/j.jsb.2016.03.002>
- [25] “Major axis and minor axis”, available at, <https://courses.lumenlearning.com/suny-osalgebra1trig/chapter/the-ellipse/>
- [26] Yang, Y., & Liu, X. (1999, August). A re-examination of text categorization methods. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 42-49). DOI: <https://doi.org/10.1145/312624.312647>

BIOGRAPHY OF AUTHORS



Amogha B, a research scholar in Reva University. Finished his M.Tech from BVBCET, Hubli. His area of interest includes, image processing, feature optimization, pattern recognition, computer vision, machine learning and deep learning. Currently working on the PhD topic of identification of person based on their walking style.
ORCID ID - 0000-0001-9782-5655



Dr. Rohini Deshpande received her Ph. D degree in Electronics and Communication Engg., from Guru Gobind Singh Indraprasth University (GGSIPU) New Delhi, M E degree in Electronics and Communication Engg, from the University of Delhi, with honors and Bachelor degree in Electronics and Communication Engg, from Gulbarga Univeristy. She is in academics for the last 32 years and involved in research and teaching UG and PG students. Her major areas of research are Signal Processing, Optical Networks, Mobile Computation etc. She has published more than 36 papers in journals and conferences of good impact factors. She has citation index of 69 and h-index of 5. She has given invited talk in various occasions including many international conferences. She also has administrative experience of around 10 years. She has delivered the responsibilities of head of ECE, IT, CSE and Mechatronics Engineering departments. She has research experience of 17 years. She has guided one scholar and is guiding five scholars for their doctoral degrees. Presently she is professor in school of ECE, REVA University, Bengaluru. She can be contacted at email: drrohini.deshpande@gmail.com
Vidwan ID: 231563/<http://vidwan.inflibnet.ac.in/profile231563>
Google scholar ID: [uulD9BEAAAAJ](https://scholar.google.com/citations?hl=en&user=uulD9BEAAAAJ)
Scopus ID: 57197052364,
ORCID ID: 0000-0002-2419.5223