Multimedia Tools and Applications (2022) 81:43919–43945
https://doi.org/10.1007/s11042-022-13252-w

Ear recognition with ensemble classifiers; A deep learning approach

Maha Sharkas [1,2]

Received: 24 January 2021 / Revised: 18 January 2022 / Accepted: 16 May 2022/
Published online: 30 May 2022

Abstract
Biometrics has emerged as a major domain for security systems. Ear as a biometric has many distinctive features which makes it promising for personal identification systems. In this paper, two tracks for classification of ear images are implemented and tested. The first employs a classical machine learning technique based on extracting features from the discrete curvelet transform and passing the extracted features to a classifier. Image preprocessing is needed for enhancement and segmentation. Ear region is first selected from the background then the curvelet transform via wrapping is applied on the segmented ear images. Different levels are investigated. The coarse image is divided into blocks and the mean, variance and entropy are calculated for each block and concatenated with the same calculated statistical features from the subimages at different levels forming the feature vector. The feature vector is passed to a classifier for ear recognition and the only classifier that provided comparative results was the ensemble classifiers. In the second track, deep learning methods are employed. Different end-to-end networks are used for classifying ear images. Features are then extracted from each network and fed to a shallow classifier for ear classification. Principal component analysis is used for feature reduction. Different classifiers are again investigated and the only classifiers which succeeded to give superior results are the Ensemble classifiers. The achieved classification rate showed improved results compared to the published methods that proves the superiority of the Ensemble classifiers for correctly classifying ear images.

Keywords Biometrics · Ear recognition, · Ensemble classifiers · Deep learning

✉ Maha Sharkas

[1] Electronics and Communications Engineering Department, Arab Academy for Science & Technology, Alexandria, Egypt

[2] Abu Kir, Egypt

1 Introduction

Personal identification systems based on ear recognition is an active research area in biometrics. Ear images are captured from a distance which makes the technology an appealing choice for surveillance and security applications as well as other application domains.Unlike faces, ears are relatively constant over a person's life and are unaffected by expressions, which make them a particularly appealing approach to noncontact biometrics [8].

The ear structure is rich and stable and is permanent over the human life and is quite unique in individuals. Also, it is invariable to the changes in pose and facial expression. Furthermore, it is relatively immune from anxiety, privacy, and hygiene problems like several other biometric candidates. Therefore, automated personal identification systems using ear images have been studied intensively for possible commercial applications [24]. Ear as a Biometric has some advantages over other biometrics like iris, fingerprints, face and retinal scans in that it is large compared to iris and fingerprint and image acquisition of the human ear is very simple and can be captured from a distance.

The anatomy of human ear is given in Figure 1. The human ear is an extremely arched 3D surface which has 3D discriminant features for human identification and recognition. Figure 1 shows various parts of the outer ear image such as helix, fossa, crus-antihelix, anti-helical fold, lower antihelix, antitragus, tragus and upper & lower concha.

It has been found that no two ears are exactly the same even that of identical twins [4] and [26].

For all the previous reasons, Ear biometrics can be used for computerized human identification and verification systems. One of the major applications of this technology is crime investigation and forensic sciences for recognition.

In this paper, two scenarios for ear classification are implemented. The first used the Discrete curvelet transform (DCT) which was especially designed to link scale with

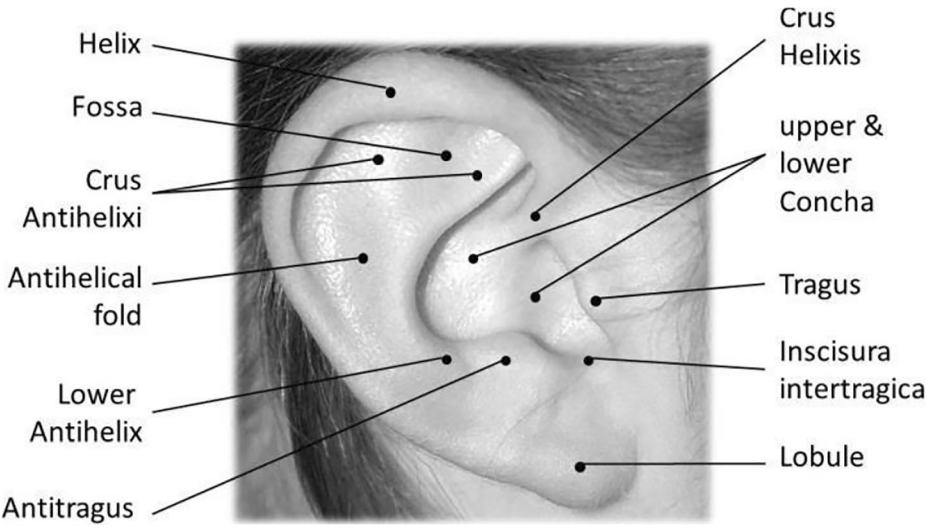

Fig. 1 Ear image

orientation. The DCT performs a multiscale and multidirectional expansions that provides good representation for objects with edges especially for objects which are smooth except for discontinuities along general curves with bounded curvature. The DCT was previously used for ear recognition in [1] with ear database from IIT Delhi ear databases. No segmentation method was used as the ear images were already cropped. A feature vector of the coarsest level image and the maximum coefficient of each image at the second coarsest level at eight different angles is generated. The K- nearest neighbor was used for classification. The recognition rate reached 97%.

In this paper features were extracted from the DCT in a different manner, which will be explained in later sections, from the AMI ear database which is not previously segmented and the ensemble classifiers were used for classification.

The second scenario employs deep learning. Here, three pretrained networks namely: AlexNet, GoogleNet and ResNet50 performed end-to-end ear classification. Two databases are investigated, the AMI ear database with 100 classes and IIT Delhi ear database with 125 classes. Classification accuracies are measured and compared.

Features from selected layers in the three networks are extracted and again passed to the ensemble classifiers for classification. The Principal component analysis PCA with different variance levels is used for feature reduction. Classification results at different levels are compared. A block diagram of the proposed system is shown in Fig. 2.

The rest of the paper is organized as follows: In Section 2, a background on different methods for ear recognition is introduced. The details of the first scenario for ear classification is given in Section 3 followed by its results in Section 4. Deep learning is briefly introduced in Section 5 and the second scenario is explained. Results of second scenario are given in Section 6. A discussion of the results is provided in Section 7 and finally the paper is concluded in Section 8.

Research contributions can be summarized as follows:

1. Designing two tracks for personal identification systems based on ear recognition.
2. The first track implements a classical machine learning method which goes through segmentation, feature extraction using the DCT and classification using Ensemble classifiers.
3. The other track investigates deep learning methods using different CNNs.
4. The results are compared with the latest state-of-the-art methods using the same datasets which proved the superiority of the proposed algorithms.

2 Background

The field of Biometric identification systems usually explores two tracks one using different feature extraction and classification methods, which are the traditional machine learning methods and the other track involves deep learning methods.

We will start by investigating some research based on machine learning techniques which started with Burge and Burger in 1998 with the first automatic ear recognition technique which was based on an adjacency graph built from Voronoi regions of ear-curve [7]. In 1999 Moreno et al. [27] presented the first fully automated ear recognition procedure using geometric

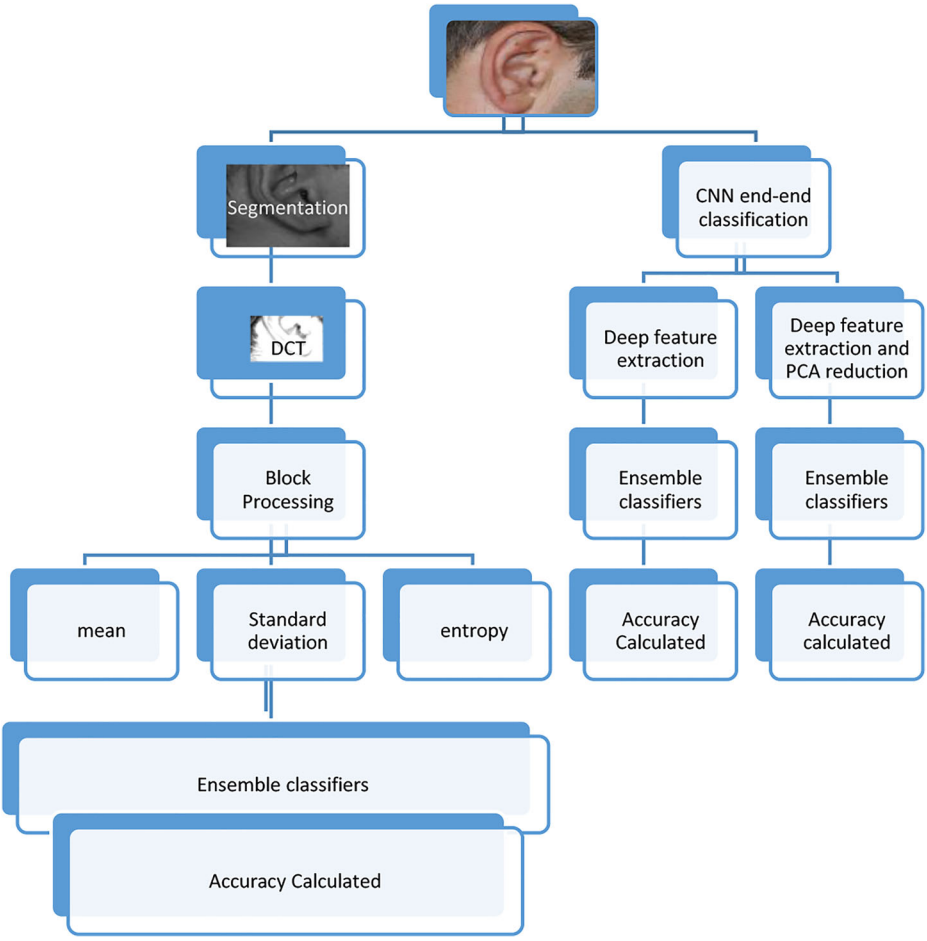

Fig. 2 Block diagram of the proposed system

characteristics of the ear and a compression network. In 2000, an ear recognition technique based on the Force Field Transform was suggested in [16].

The Forensic Ear Identification Project (FEARID) project was launched in 2001, marking the first large-scale project in the field of ear recognition [19]. In this project, ear prints are studied to investigate their strength for crime scenes. Three left and three right ear prints were collected from three countries. The equal error rate was used for evaluation which was 4% for lab quality images but increased to 9% for print vs mark comparisons.

Later several ear recognition techniques were implemented. Victor et al. [5] applied principal component analysis (PCA) on ear images which gave promising results but results proved that face is a more reliable biometric that the ear.

The field force transform was used in [17]. The method was implemented on 252 ear images taken from 63 subjects from the XM2VTS face database. The accuracy reached 99.2 for poorly registered and extracted ear images and dropped to 62.4 when using PCA but then increased to 98.4 with accurate extraction and registration. In 2006 a method based on non-negative matrix factorization (NMF) was developed by Yuan et al. [39] and was applied to

occluded and non-occluded ear images from the USTB ear data base. Ears are manually extracted and three ears are used for training and the fourth for testing. The best recognition rate reached 91%. The drawbacks hers are manual extraction. A method based on the 2D wavelet transform was introduced by Nosrati et al. [28] in 2007, followed by a technique based on log-Gabor wavelets in the same year [23]. In [28] the authors used the 2D wavelet transform for feature extraction on two databases which are the USTB and Carreira-Perpinan. Accuracies reached 90.5% for the USTB database for two images out of four for training and in the case of the Carreira-Perpinan dataset accuracies reached 95.05 for three images out of four for training and 97.05 for four images for training. Here the accuracy increased when more images are used for training and in the case of four images, all images are training images.

In 2011, local binary patterns (LBP) was used for ear image description in [38]. Binarized statistical image features (BSIF) and local phase quantization (LPQ) features were used and their results are given in [6, 30, 31]. In [6] the authors used the Binarized Statistical Image Features to extract features from three databases IIT Delhi 1 & 2 and USTB database. The results reached 97.26, 97.34 and 98.46 respectively with KNN classifiers. The authors in [30] used a combined database from three data sets. The used dataset has 2432 images from 555 subjects which are: 363 subjects from UND-J2, 67 subjects from AMI and 125 subjects from IIT Delhi with at least two samples per subject. The hit rate reached 96.89 with the use of PCA which is used for dimensionality reduction and improved to 99.01 when multi-cluster search strategy is used. The authors continued their work in [31] where the best results were achieved for three different datasets when the LPQ and the BSIF descriptor were combined for feature extraction, LDA is used for dimensionality reduction and the cosine distance for classification. The authors in [24] proposed an automated human identification using 2D ear imaging. They presented a segmentation method based on morphological operations and Fourier descriptors. They extracted ear features using localized orientation information and examined local gray level phase information using complex Gabor filters. The rank-one recognition accuracy reached 96.27% and 95.93%, respectively, on the database of 125 and 221 subjects from the IIT-Delhi database.

In 2014 the first ear recognition system based on curvelet features was published in [1]. The feature vector of each image is composed of the approximate curvelet coefficients and maximum coefficients of second coarsest level curvelet coefficients at eight different angles. The k-NN (k-nearest neighbor) is utilized as a classifier. The accuracy reached 97% for the IIT-Delhi database. Here the author used segmented ear database from the IIT-Delhi database.

The second track, which is based on deep learning methods is investigated in the following research papers.

In [18], the authors fused deep features from different layers using discriminant correlation analysis and used pairwise SVM and KNN for classification. They applied their work on the USTB I and II and IIT-Delhi I and II ear databases and achieved an accuracy rate of above 99%. In [40], the authors proposed a new ear database under uncontrolled condition and tested the classification accuracy with CNN. They changed the last pooling layers with spatial pyramid pooling (SPP) layers in order to fit arbitrary data size and obtain multi-level features. They achieved a max accuracy of about 97%.

The authors in [11] proposed a deep learning model for unconstrained ear recognition. They passed the features to a shallow classifier for ear recognition. They suggested a deep learning–

based averaging ensemble to limit the over fitting with best achieved results with an ensemble of ResNet18 models which provided consistent performance across the tested datasets. In [12], the authors created a new ear dataset and called it multi-pie ear dataset. The classification accuracy was improved by combining the output of different CNN models. In [13], the authors proposed a fusion of learned deep features with handcrafted features for unconstrained ear recognition. They reached a conclusion that handcrafted features are not dead and they improve the performance.

Scorenet, which is a deep cascade level fusion is proposed in [20]. Here, the authors fuse deep features from different levels of different CNN networks with handcrafted features for unconstrained ear recognition.

In [29], the authors created an earcode from the first principal component obtained by Kernel PCA. The authors created their own dataset from 103 persons. The performance rates were comparatively high with EER of 0.13 and TPR of 0.85. Also, the authors applied the algorithm on standard databases which are IIT1 and USTB1 databases and achieved comparative results.

The authors in [2] created a human recognition algorithm based on fusion of ear and tragus in a single image to overcome the challenges of partial occlusions, pose variations and week illumination. The Local binary pattern is used for feature extraction with score-level fusion and KNN used for classification. The experiments were implemented on USTB 1,2 and 3 dataset and gave comparative results.

The authors in [21] discussed the lack of color information in ear images and its effect on the accuracy. They suggested a framework responsible for colorizing grayscale and dark images followed by a classification task. The algorithm is implemented on two databases which are the constrained AMI and the unconstrained AWE ear datasets and provided an accuracy of 96 and 50.53 respectively.

In [3], the authors presented an ear recognition system based on CNN especially VGG networks. The best models were used to build ensembles of models with varying depth. The work was implemented on the AMI and WPUT ear datasets and also the AMIC Database which is the original AMI database but with critically cropped background. The rank1 classification accuracy reached 97.5 for VGG ensembles of 13-16-19, and 93.21 for AMIC with VGG ensembles of 11-13-16-19 and 79.08 for the WPUT database for the same VGG ensembles.

It is noticed here that the recognition accuracy is reduced with AMIC which is a segmented version of the AMI database and this is due that cropping profile images may result in losing important information.

The local binary pattern LBP and its use to extract features from ear images is discussed by the authors in [14]. They investigated its performance over five benchmark databases which are the IIT Delhi I and II, AMI, WPUT and AWE. The results showed good performance in case of constrained images while the accuracies decreased a lot with increased distortions.

A six later deep CNN was proposed by the authors in [33] and was tested on IIT Delhi II and AMI ear datasets with a recognition rate accuracy that reached 97.36% and 96.99% respectively for 1000 epochs. The results are repeated on the AMI dataset where the ear images are rotated with different angles with different illumination conditions and also when adding random noise. The recognition accuracy decreased and reached 91.99 for the combined variation conditions.

3 Ear classification with DCT features

In this section, the framework of the first scenario is introduced. The ear recognition algorithm starts with a simple segmentation method based on filtering and morphological operations. The segmentation method cropped the ear images leaving a part of the background. The discrete curvelet transform via wrapping is employed for feature extraction. Statistical features are extracted from the curvelet images at different levels. Different levels are investigated (three levels, four levels and five levels). The coarsest level image is divided into blocks and the mean and standard deviation are calculated for each block and concatenated with the same features extracted from images at different fine levels forming the ear feature vector. The entropy is then added using the same technique forming a new feature vector. The ensemble classifier using subspace discriminant analysis is used for classification.

The proposed algorithm is investigated in the following sections.

3.1 Ear segmentation

The AMI Ear Database used in this first scenario was created by Esther Gonzalez during her work on the PhD in Computer Science. The ear database contains images which are collected from students, teachers and staff of the Computer Science department at Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas, Spain. The images are captured in an indoor environment. The database was collected from 100 different subjects in the age group from 19–65 years. Seven images (six right ear images and one left ear image) were taken for each individual.

Samples of the used database are shown in Fig. 3.

Nikon D100 camera was used to capture all the images under the same lighting conditions, with the subject placed seated at a distance of about 2 meters from the camera and looking at

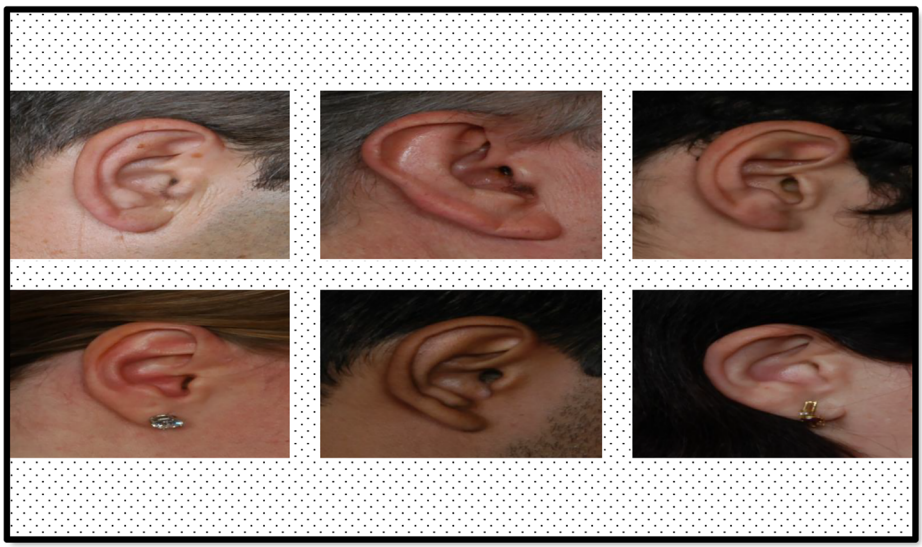

Fig. 3 Sample images from the used database

some previously fixed marks. Six of the seven images used A135 mm focal length while the 200 mm focal length was used for the image that was called ZOOM. From the captured images, five of them were right side profile images (right ear) with the subject facing forward (FRONT), looking up and down (UP, DOWN) and looking left and right (LEFT, RIGHT). The sixth image of right profile was taken with the subject also facing forward but with a different camera focal length (ZOOM). The last image which is the (BACK) image was a left side profile (left ear) and the subject in this case is facing forward and the same camera focal length is used as the previous five images.

The database consists of 700 images and has been sequentially numbered for every subject with an integer identification number. Images have a resolution of 492 x 702 pixels and are available in jpeg format.

94 subjects each having six images excluding the back ear comprising 564 images are used in this paper. Six subjects and the back ear are removed from the database in the first scenario because they were not correctly segmented.

All images are converted from color images to greyscale images where grayscale values are formed by forming a weighted sum of the *R (red)*, *G (green)*, and *B (blue)* components as follows:

$$0.2989 * \mathrm{R} + 0.5870 * \mathrm{G} + 0.1140 * \mathrm{B}$$

Image segmentation starts with low pass filtering the ear image followed by applying a grey level threshold to convert to a binary image then in-between holes are filled. The threshold chosen is a global threshold using Otsu's method. Otsu's method chooses a threshold that minimizes the intraclass variance of the thresholded black and white pixels and is used for binarization. Otsu's thresholding method involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels at each side of the threshold, i.e. the pixels that either fall in foreground or background, then in-between holes are filled.

All connected components (objects) that have fewer than P pixels are removed, producing another binary image. The value of P is experimentally chosen to be 150.

Mapping is performed between the processed binary image and the original image to produce the final segmented ear image.

Ear segmentation Steps are demonstrated in Fig. 4.

It can be noticed that a very simple segmentation method is used and a part of the background is included.

Samples of other segmented ear images are shown in Fig. 5.

3.2 The discrete curvelet transform

The curvelet transform was suggested by E. J. Cand`es and D. L. Donoho in [9]. The Curvelet transform is a geometric transform created to overcome the limitations of wavelet like transforms. Curvelet transform is a multi-scale and multi-directional transform with needle shaped basis functions. The basis functions of the wavelet transform are isotropic therefore, it requires large number of coefficients to represent the curve singularities. On the other hand, the basis functions of the Curvelet transform are needle shaped and have high directional

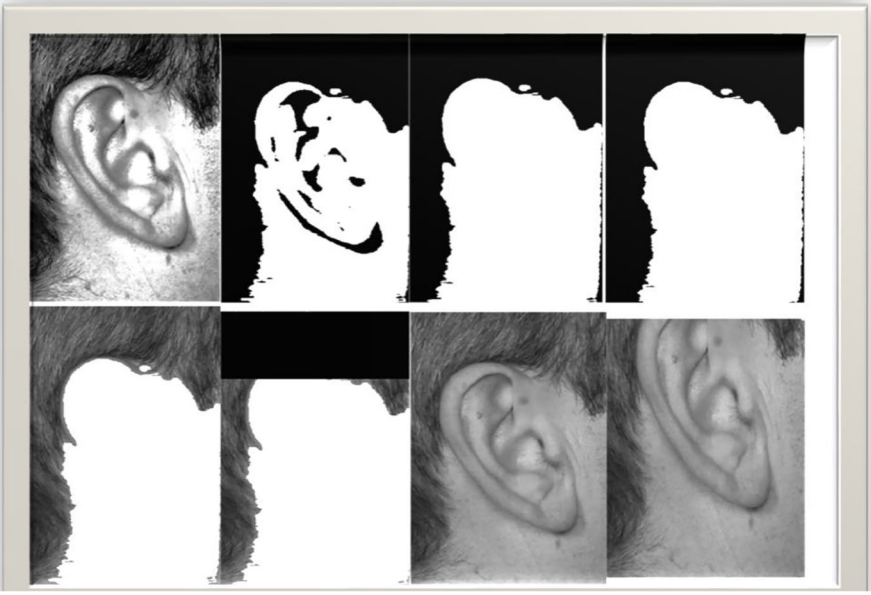

Fig. 4 The segmentation process

sensitivity and anisotropy. Also, they obey parabolic scaling and therefore the Curvelet transform allows almost optimal sparse representation of curve singularities.

The Curvelet transform was designed to represent edges and other singularities along curves much more efficiently than traditional transforms by using fewer coefficients for a given accuracy of reconstruction.

The origin of the Curvelet transform is a ridge transform added with a binary square window. However, there exists big data redundancy in the transform. Therefore, the first generation curvelet transform can be improved to obtain the 2nd generation, and the second takes on features with faster computation and less redundancy. Curvelets as a function of x=(x1, x2) at scale 2^{-j}, orientation θ_t and position $x_k^{(j,l)} = R_{\theta^l}^{-1}\left(k_1 2^{-j}, k_2, 2^{-\frac{j}{2}}\right)$ are defined as [10]

$$\phi_{j,l,k}(x) = \phi_j\left(R_{\vartheta_i}\left(x - x_k^{(j,l)}\right)\right) \tag{1}$$

R_θ is the rotation in radians and is given by

$$R_\theta = \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} \tag{2}$$

and R_θ^{-1} is its inverse

To calculate the curvelet coefficient then apply the inner product between an element $f \in L^2(R^2)$ and a curvelet $\phi_{j,l,k}(x)$

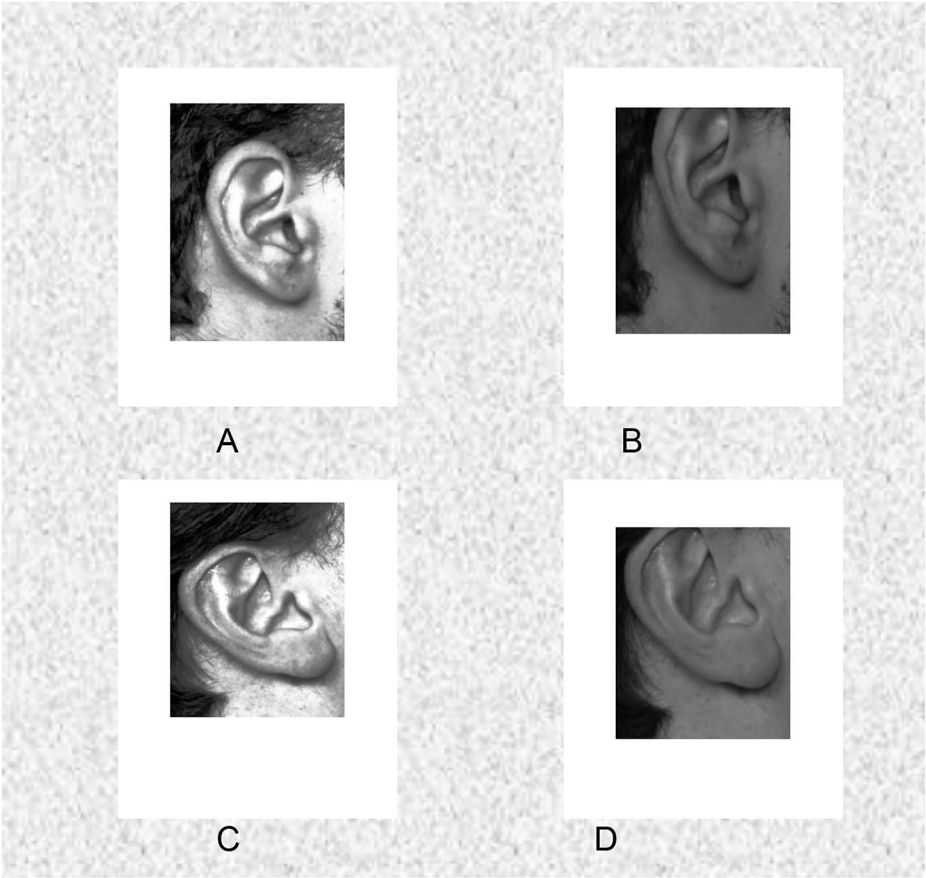

Fig. 5 Samples of segmented ear images. **A** Original image 1, **B** Segmented image 1, **C** Original image 2, **D** Segmented image 2

$$c(j,l,k) := \left(f, \phi_{j,l,k}\right) = \int_{R^2} f(x)\overline{\phi_{j,l,k}(x)}dx \tag{3}$$

In the digital form the digital curvelet transform is given by

$$C^D(j,l,k) = \sum\nolimits_{0 \le t_1, t_2} \tag{4}$$

Where $f(t_1, t_2)$ is an input cartesian array and $o \le t_1, t_2$

The notation D stands for digital, $\phi^D{}_{j,l,k}$ is the digital curvelet transform and $C^D(j,l,k)$ is a collection of coefficients.

The digital curvelet transform is implemented using fast discrete curvelet transform. It is computed in the spectral domain to use the advantage of FFT. The image and the curvelet are both transformed to the Fourier domain and the curvelet is convolved with the image in the

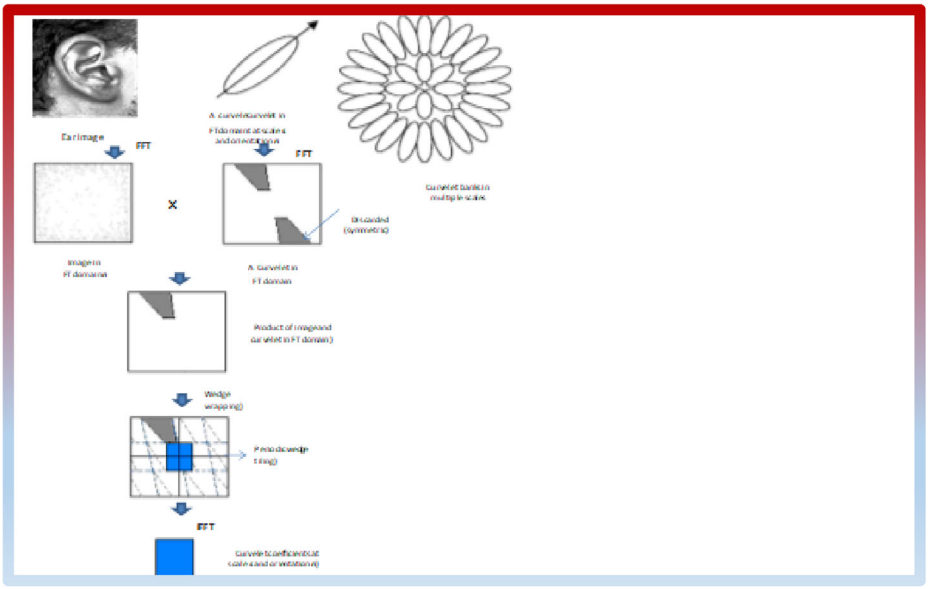

Fig. 6 Feature extraction using the curvelet transform [2]

spatial domain which becomes the product in the Fourier domain. The curvelet coefficients are finally obtained by applying the inverse Fourier transform on the spectral product.

The frequency response of a curvelet is a non-rectangular wedge, and this wedge is therefore needed to be wrapped into a rectangle to be able to apply the inverse Fourier transform. The wrapping is performed by periodic tiling of the spectrum using the wedge, and then the rectangular coefficients area is collected in center. The rectangular region collects the wedge's corresponding portions from the surrounding periodic wedges using this periodic tiling [36]. The complete feature extraction process using a single curvelet is illustrated in Fig. 6.

In this paper, the ear images are segmented and then transformed using the discrete curvelet transform DCT via wrapping into the curvelet domain. The DCT is a redundant transform. The DCT is implemented in different decomposition levels (three levels, four levels and five levels) and feature vectors are created.

In the three level decomposition, we have the coarsest level and one fine level at eight different angle. An example of the decomposed ear image is shown in Fig. 7.

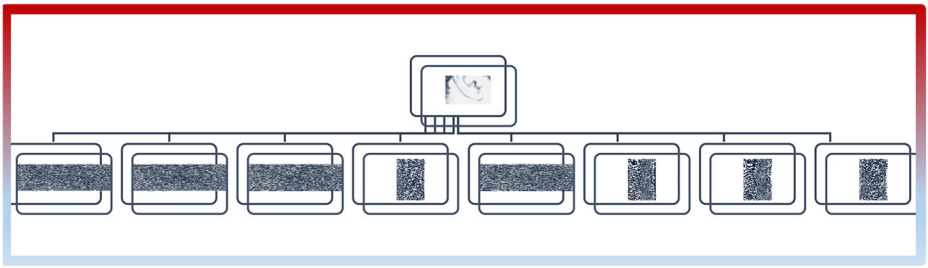

Fig. 7 Three level decomposition of the ear image at the coarsest level and the fine level at eight different angles

Table 1 Accuracy results for feature vectors of mean and variance

Number of Levels	Length of Feature Vector	Accuracy Results (%)
3 Levels	356	70.6
4 Levels	148	77.8
5 Levels	180	62.1

At the three levels decomposition, the coarsest level image is of size 85×85 and is divided into blocks of size 7×7. In each block, the mean (mn) and standard deviation (stnd) are calculated. We have 169 blocks resulting in 169 values for mn and the same for stnd. Also, the mn and stnd are calculated for all image in the fine level, then again we get 8 values for mn and the same for stnd. Finally, the last level has one value for mn and same for stnd. All these values are concatenated to get a feature vector of length [169 + 169 + 8 + 8 + 1 + 1 = 356].

The same procedure is done for the four level decomposition with a coarsest level image of size 43×43 and 2 fine levels. The first level has images at eight different angles and the second fine levels has images at sixteen different angles.

The feature vector is of size [49 + 49 + 8 + 8 + 16 + 16 + 1 + 1 = 148].

Again, for the five level decomposition we have a coarsest level image of size 21×21 and three fine level at sixteen, thirty-two and thirty-two images at different angles respectively.

The feature vector is of length 180.

The classification results using the ensemble classifier will be shown later.

Another statistical parameter is added which is the entropy, and the same procedure is implemented. The entropy (E) is a statistical measure that measures the randomness and is used to estimate the texture in an input image and can also measure the distribution variation in a region. The feature vector length became 534 for the three levels, 222 for the four levels and 270 for the five levels. The mathematical equations for the mean, standard deviation and entropy in each block Z are given below.

$$mn_z = \frac{1}{FG}\sum_{i,j=1}^{FG} p_{z(i,j)} \tag{5}$$

Where $p_{z(i,j)}$ are the pixels in each Z block and F×G is the block size.

Table 2 Accuracy results for feature vectors of mean, variance and entropy

Number of Levels	Length of Feature Vector	Accuracy Results (%)
3 Levels	534	78.9
4 Levels	222	86.3
5 Levels	270	67.9

Table 3 Accuracy results for different block sizes for the 4-levels configuration

Block size	Length of Feature Vector	Accuracy Results(%)
6×6	267	85.6
8×8	183	85.6
9×9	150	86.5
10×10	123	85.3
13×13	102	85.8
15×15	102	82.6
17×17	102	82.8
Full Block	78	72

$$stnd_z = \sqrt{\frac{1}{FG}\sum_{i,j=1}^{FG}(p_z(i,j) - mn_z)^2} \tag{6}$$

$$E_z = -\sum_{i=0}^{n-1} pr_i \times lnpr_i \tag{7}$$

where n is the number of grey levels and pr_i is the probability of a pixel having gray level i.

Again, the classification results will be discussed later.

3.3 The ensemble classifier

Ensemble classifier are a group of individual classifiers that are cooperatively trained on data sets to solve a supervised classification problem [34].

The base classifiers are trained separately on the data set to give a decision on a test pattern. The decisions arethen combined by a suitable fusion method. A number of fusion methods are discussed in the literature and include majority voting, Borda count, algebraic combiners etc. [32].

Several classifiers were investigated for ear classification using the features obtained in the first scenario including decision trees, supervised vector machine SVM, K nearest neighbors and ensemble classifiers which were the ones that provided the best results especially with the subspace discriminant ensemble classifier. The number of learners are chosen to be 30, and the subspace dimension is set to be 178. The classifier has 94 input classes and uses a 5-fold cross validation.

Table 4 Accuracy results for feature vectors of mean, variance and entropy (9 × 9) block

Number of Levels	Length of Feature Vector	Accuracy Results (%)
3 Levels	327	80.9
4 Levels	150	86.5
5 Levels	270	75.4

Table 5 t test results for the (9 × 9 block)

t-score	SED	DOF	MD	CR	P value
848.3983	0.102	9	86.52	86.28–86.75	<0.05

4 Results of the first scenario

As mentioned earlier, the AMI ear image database is used in this scenario. 94 subjects each having six images comprising a total of 564 images are the database used. The DCT via wrapping is implemented at different decomposition levels (3, 4 and 5 levels).

The curvelet transform is a completely redundant transform therefore features have to be carefully selected. Several experiments are implemented and tested. The coarsest level image is divided into sub-blocks of size 7×7, and the mean and variance are calculated for each block and concatenated with the calculated mean and variance of the subimages at the fine levels forming the feature vector.

For the three level decomposition, the coursest level image is of size 85×85 and the second coarsest level have subimages at eight different angles. Similarly, for the four levels, the coarsest level image is of size 43×43 and the finest levels have eight then sixteen different angles respectively. For the five levels, the coursest level image is of size 21×21 with sixteen, thirty-two and thirty-two angles at the second, third and fourth finest levels respectively.

Several classifiers were used for training and testing including Supervised Vector Machine (SVM), K nearest neighbor (KNN), Forest trees and Ensemble Classifiers which proved to produce the best accuracy results. The Ensemble classifier is used to test the accuracy. The accuracy is given by:

$$accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \tag{8}$$

where t_p is the true positive rate, t_n is the true negative rate, f_p is the false positive rate and f_n is the false negative rate.

The accuracy results are shown in Table 1.

As noticed from Table 1, the best achieved accuracy was when using four levels which reached 77.8%.

Another feature is added in an attempt to improve the accuracy. The entropy is calculated in the same way as the mean and variance and concatenated with the previously calculated feature vector resulting in a new feature vector with three components; mean, variance and entropy. The accuracy results using the ensemble classifier are tabulated in Table 2.

Table 6 Accuracy results for feature vectors of mean, variance and entropy from raw images (7 × 7 block)

Number of Levels	Length of Feature Vector	Accuracy Results (%)
3 Levels	534	79.1
4 Levels	222	83.3
5 Levels	270	66

Table 7 Accuracy results for feature vectors of mean, variance and entropy from raw images (9 × 9 block)

Number of Levels	Length of Feature Vector	Accuracy Results (%)
3 Levels	327	80.9
4 Levels	150	82.8
5 Levels	270	64.5

The number of blocks in the coarsest level image are selected to be of size 7×7. It is clear from the accuracy results that the best configuration is the 4 levels so this configuration is taken as the model for all the coming experiments. Several other block sizes are investigated and their results are summarized in Table 3.

As noticed in Table 3, the achieved accuracy when dividing the coarsest level image into 9×9 blocks has almost the same accuracy as 7×7 block bit with a smaller feature vector (150 coefficients instead of 222 coefficients) which provides less computational complexity.

To evaluate the 9×9 block for all levels, the same procedure is done and the results are tabulated in Table 4.

Table 4 indicates that in spite of that the feature vector decreased in length, the accuracy increased.

Results for the (9 × 9 block) are repeated 10 times and validated by performing a t test. The results are given in Table 5 and resulted in a p value < 0.05.

Although the segmentation algorithm used here is very simple, the ensemble classifiers are able to provide competitive results which can be used for medium security applications.

The question is, is the segmentation process necessary; that is if the curvelet features are extracted directly from the raw ear images which consequently reduces the processing time, how will be the accuracy affected.

Table 8 Alexnet layers

Layer Label	Specifications		Output Dimension
Input layer			227×227×3
Convolution Layer 1	Filter Size	11×11	55×55×96
	Stride	4	
	Padding	0	
Pooling Layer 1	Pooling Size	3×3	27×27×96
	Stride	2	
Convolution Layer 2	Filter Size	5×5	27×27×256
	Stride	1	
Pooling Layer 2	Pooling Size	3×3	13×13×256
	Stride	2	
Convolution Layer 3	Filter Size	3×3	13×13×384
	Stride	1	
Convolution 4	Filter Size	3×3	13×13×384
	Stride	1	
Convolution Layer5	Filter Size	3×3	13×13×256
	Stride	1	
Pooling Layer 5	Pooling Size	3×3	6×6×256
	Stride	2	
FC6 Layer			4096×2
FC7 Layer			4096×2
FC8 Layer			1000×2

Table 9 Googlenet layers

Layer Name	Filter Size	Stride	Output Size
Input Layer			224×224×3
conv1	7×7	2	112×112×64
pool1	3×3	2	56×56×64
conv2	3×3	1	56×56×192
pool2	3×3	2	28×28×192
Inception (3a)	-	-	28×28×256
Inception (3b)	-	-	28×28×480
pool3	3×3	2	14×14×480
Inception (4a)	-	-	14×14×512
Inception (4b)	-	-	14×14×512
Inception (4c)	-	-	14×14×512
Inception (4d)	-	-	14×14×528
Inception (4e)	-	-	14×14×832
pool4	3×3	2	7×7×832
Inception (5a)	-	-	7×7×832
Inception (5b)	-	-	7×7×1024
average pooling	7×7	1	1×1×1024
fully connected (fc)			1024×2

The accuracy results for the 7×7 coarsest level block division and 9×9 coarsest level block division for raw images without segmentation are tabulated in Tables 6 and 7 respectively.

Again, the best achieved accuracy result was for the four level decomposition.

Table 10 Resnet50 layers

Layer Label	Input Layer Dimension	Output Dimension
Input Layer		227×227×3
Conv1	112×112×64	Filter size=7×7 Number of filters=64 Stride=2 Padding=3
pool1	56×56×64	Pooling size=3×3 Stride=2
conv2_x	56×56×64	$\begin{bmatrix} 1\times1. & 64 \\ 3\times3. & 64 \\ 1\times1. & 256 \end{bmatrix} \times 3$
conv3_x	28×28×128	$\begin{bmatrix} 1\times1. & 128 \\ 3\times3. & 128 \\ 1\times1. & 512 \end{bmatrix} \times 4$
conv4_x	14×14×256	$\begin{bmatrix} 1\times1. & 256 \\ 3\times3. & 256 \\ 1\times1. & 1024 \end{bmatrix} \times 6$
conv5_x	7×7×512	$\begin{bmatrix} 1\times1. & 512 \\ 3\times3. & 512 \\ 1\times1. & 2048 \end{bmatrix} \times 3$
Average pooling		Pool size=7×7 Stride=7 1×1×2048
FC Layer		2 (2048×2)

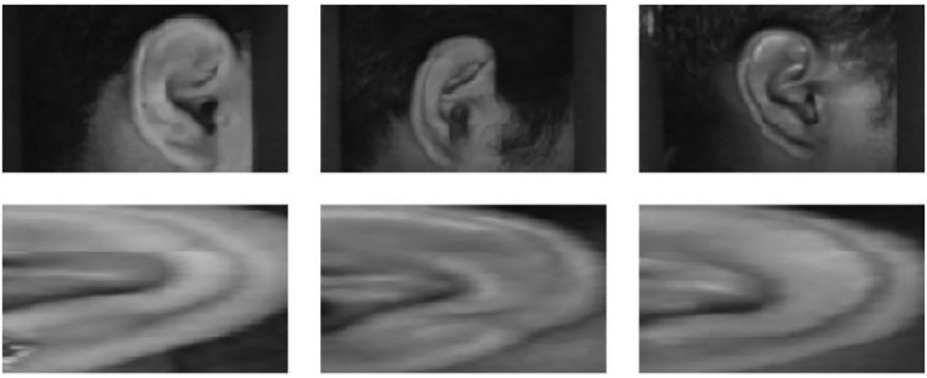

Fig. 8 Sample images from the raw and segmented IIT-Delhi ear database

The non-segmented ear recognition can be used for high speed, medium security applications.

The superiority of the proposed techniques is with the simple segmentation method with the use of the ensemble classifiers. A small feature vector with only 150 coefficients extracted from curvelet coefficients is used which reduces the computation time and yields reasonable results suitable for medium security applications. To compare the results with the state-of-the art models on the same database, we can see that the proposed methods outperforms the results obtained in [14] which provided an accuracy of 73.73 using local binary patterns.

In the next section, we will introduce the deep learning methods which produced superior results.

5 Deep learning

Progress in convolutional neural networks CNNs encouraged researchers to implement the different structures in many applications such as image classification, object detection, medical image applications, face recognition etc. Examples of some applications are given in [25] and [35]. Deep networks extract low, middle and high-level features and classifiers in an end-to-end multi-layer fashion, and the number of stacked layers can enrich the "levels" of features. Recently some research for ear recognition using deep features are done and are discussed earlier.

In this paper, deep learning is employed for ear recognition. Three well known pre-trained networks are investigated. In general, the top layers in a CNN network contain semantic information while the intermediate layers describe the local features. Low level features which describe textures and edges are in the bottom layers.

Table 11 End-to-end percentage classification accuracies

Ear database	AlexNet	GoogleNet	ResNet50
AMI	94	88.5	99
IIT Delhi (raw)	94.29	90.71	93.57
IIT Delhi (segmented)	62.86	21.43	59.29

Table 12 t test results for the Resnet50 classification accuracies with the AMI database

t-score	SED	DOF	MD	CR	P value
542.2453	0.1826	9	99	98.587–99.413	<0.05

Different scenarios are implemented. First, end-to-end ear recognition is performed through AlexNet, GoogleNet and ResNet50.

AlexNet won the The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 for an input image of 1000 classes [22]. The AlexNet architecture has five convolutional layers and three fully connected layers. Alexnet layers are given in Table 8.

GoogleNet [37] is the winner of the ILSVRC 2014 competition. Their architecture consisted of a 22-layer deep CNN but reduced the number of parameters from 60 million in case of AlexNet to 4 million.

As the network gets deeper and starts to converge, sometimes the performance degrades as the network saturates. This is not due to overfitting or adding more layers but due that not all systems are easily optimized.

Layers of GoogleNet are shown in Table 9.

ResNet [15] was able to overcome this problem by introducing shortcut connections which skip one or more layers. These connections do identity mapping by adding their outputs to the outputs of the stacked layers. Resnet50 layers are given in Table 10.

To overcome the limited data size which can cause overfitting, data augmentation is implemented. This process generates batches of new images from the original data with some preprocessing such as resizing, rotation, translation and reflection.

Transfer learning is a technique commonly used in deep learning where a model trained on a certain task is used for another task. Transfer learning is an optimization that allows rapid progress or improved performance when modelling the second task. The network is first trained on a dataset, and then the learned features are transferred to a second target network to be trained on a target dataset and task. This process works well if the features are general, which means that the features are suitable for both the base and the target tasks, instead of specific to the base task.

In the second scenario, deep learning is used for ear recognition. Two ear databases were investigated, the AMI ear database which was investigated in the first scenario, but with 100 classes each having seven images with a total of 700 images not 94 as in the first scenario. The second used database is the IIT Delhi with 125 classes each having at least three images with a total of 493 images. IIT Delhi database has a database for raw ear images and another one for segmented ear images, which are both investigated while, for the AMI ear database only raw

Table 13 Percentage classification results for features of AlexNet with ensemble classifiers on AMI database

# of Input Features	Bagged trees	Subspace discriminant	Subspace KNN
4096	79.3	98.1	91.6
136 PCA with 95% var	64.3	97	91.6
207 PCA with 97% var	61.3	97.3	91.6
271 PCA with 99% var	60.6	96.4	91.9

Table 14 Percentage classification results for features of GoogleNet with ensemble classifiers on AMI database

# of Input Features	Ensemble bagged	Subspace discriminant	Subspace kNN
1024	63.7	93.3	85.9
144 PCA with 95% var	47.9	93.3	84.3
202 PCA with 97% var	52.6	93.9	84.4
336 PCA with 99% var	44.1	91.6	84.9

images are investigated. The IIT Delhi ear database is collected from students and staff at IIT Delhi, Delhi, India. Images are acquired at a distance with a simple setup and each subject has at least three images with resolution 272 × 204 pixels for raw images and 50 × 180 pixels for segmented images. Samples from IIT-Dehi raw and segmented database are given in Fig. 8, where the upper raw shows the raw images and the lower one for segmented images. End-to End deep learning using three pre-trained deep nets namely: AlexNet, GooglenNet and ResNet50, is implemented. Features are extracted from a suitable feature level of each network, which is layer fc7 in Alexnet, dropout pool5-drop_7 × 7_s1 in Googlenet and avg_pool in Resnet50, and are passed to several classifiers including decision trees, KNNs, SVMs and ensemble classifiers. PCA is used to decrease the feature vector with different variance levels. Again, the only classifiers which succeeded to give good results were the Ensemble classifiers especially the subspace discriminant. All other classifiers failed to give acceptable results. Deep learning results are given in the next section.

6 Results of the second scenario

As mentioned in the previous section, end-to-end deep learning with three deepnets are used for ear classification. Image reflection across the x-axis and image translation up to 30 pixels across and the x and y-axis are used for image augmentation. Transfer learning is applied to fine tune the used networks to the required number of classes. All experiments are done with Matlab 2018 with NVIDIA GeForce GTX 1050. All networks were trained with stochastic gradient descent with momentum, a mini batch size of 10 observations for each iteration, a maximum number of epochs equal to 20 and an initial learning rate of 0.0001.

Table 15 Percentage classification results for features of ResNet50 with ensemble classifiers on AMI database

# of Input Features	Bagged trees	Subspace discriminant	Subspace KNN
2048	60	99.45	74.3
168 PCA with 95% var	47.4	96.3	72.4
239 PCA with 97% var	49.9	96.4	72.4
393 PCA with 99% var	42	94.3	72.3

Table 16 t test results for the ResNet features passed to ensemble classifiers for AMI database

t-score	SED	DOF	MD	CR	P value
2911.5957	0.0342	9	99.45	99.3727–99.5273	<0.05

Table 11 gives the end-to-end classification accuracies for AlexNet, GoogleNet and ResNet50 for AMI and IIT Delhi raw and segmented ear images.

As clear from the previous table, that the best achieved results for the AMI database was for Resnet50 with an average mean of 99% accuracy. For the IIT Delhi raw image database, the AlexNet and ResNet50 are almost the same with 94.29% for AlexNet and 93.57 for ResNet50. A great degradation in the performance was noticed for the segmented ear image database which was the same conclusion reached before that maybe the segmentation step is not necessary.

To validate the achieved results, the best results, which is the Resnet50 with the AMI database, are repeated 10 times and a t test is performed which provided a p value less than 0.05. Results are given in Table 12.

Features are extracted from the different networks and passed to shallow classifiers for ear classification. 4096 features are extracted from AlexNet, 1024 features from GoogleNet and 2048 features from ResNet50. Several classifiers were investigated including decision trees, SVMs, KNNs and ensemble classifiers. The only classifiers which succeeded to give acceptable results were the ensemble classifiers. In fact, the subspace discriminant analysis provided superior results in most cases. PCA at different variance levels was used for feature reduction. The variance levels investigated were 95%, 97% and 99%. Again, these experiments are done on the AMI and IIT Delhi ear database for raw and segmented images.

Classification results for the different networks with the AMI database are shown in Tables 13, 14 and 15. The first column gives the number of features without PCA then with PCA at different variance levels. As noticed, the number of features were reduced to 136, 207 and 271 features at 95%, 97% and 99% of variance respectively for AlexNet features. The same can be noticed for GoogleNet and ResNet features in Tables 14 and 15.

The resnet50 features provided the best accuracy for the AMI database when passed to Ensemble classifiers subspace discriminant which achieved an average mean of 99.45%. Again the best results are repeated 10 times ans a t test is performed to validate the results which resulted in a p value less than 0.05. The details are given in Table 16.

The same procedure is done for IIT Delhi ear database for raw and segmented images. The raw images results are shown in Tables 17, 18 and 19 while the segmented images results are shown in Tables 20, 21 and 22.

Table 17 Percentage classification results for features of AlexNet with ensemble classifiers on IIT Delhi raw database

# of Input Features	Ensemble bagged	Subspace discriminant	Subspace knn
4096	55.2	93.3	90.3
117 PCA with 95% var	46	93.9	88
170 PCA with 97% var	42	91.3	89.7
294 PCA with 99% var	33.3	70.8	89.7

Table 18 Percentage classification results for features of GoogleNet with ensemble classifiers on IIT Delhi raw database

# of Input Features	Ensemble bagged	Subspace discriminant	Subspace knn
1024	58.4	93.9	87.6
105 PCA with 95% var	50.5	93.3	86.6
148 PCA with 97% var	46.2	92.3	87.2
249 PCA with 99% var	42.4	89.2	87.4

Again, best results are repeated 10 times and to validate the results a t test is performed. The best results for the IIT Delhi database is 93.9 for 117 features. These are the Alexnet features reduced with PCA at 95% variance as they provided the best results with the least number of features and the results are shown in Table 23.

A full discussion of the results is given in the next section.

7 Discussion

In this paper, two scenarios for ear classification are implemented and tested. In the first scenario, the ear image is segmented, and statistical features extracted from different levels obtained from the discrete curvelet transform are used to generate the feature vector. Statistical features are the mean, standard deviation and entropy. The DCT decomposes the ear image into a course level and fine levels. The coursest level image is divided into blocks and the mean, standard deviation and entropy are extracted for each block. The same isdone with the fine levels. Different block sizes and different fine levels are investigated with three, four and five levels with different orientations. The feature vector is then passed to different classifiers and the subspace discriminant ensemble classifier was the only classifier which succeeded to give comparative results with a classification accuracy of 86.5% for 4-level decomposition with a block size of 9×9. The length of the feature vector in this case was 150 coefficients. The author then skipped the segmentation process and passed the raw ear images directly to the DCT. The same process is done to obtain the feature vector which is then passed to the subspace discriminant ensemble classifier to obtain the classification accuracy which reached 83.3% and 82.8% for 4-level decomposition with a block size of 7×7 and 9×9 respectively.

Table 19 Percentage classification results for features ResNet50 with ensemble classifiers on IIT Delhi raw database

# of Input Features	Ensemble bagged	Subspace discriminant	Subspace knn
2048	39.6	86.6	56.4
155 PCA with 95% var	35.5	93.5	51.9
211 PCA with 97% var	32.5	90.5	52.5
323 PCA with 99% var	27	62.1	53.1

Table 20 Percentage classification results for features of AlexNet with ensemble classifiers on IIT Delhi segmented database

# of Input Features	Ensemble bagged	Subspace discriminant	Subspace knn
4096	24.3	48.3	32.7
90 PCA with 95% var	19.7	48.5	33.1
137 PCA with 97% var	15.8	48.3	32.7
190 PCA with 99% var	14.2	43.6	33.3

The AMI ear database was used in this scenario with 94 subjects each having 6 images. Seventy percent of the data was used for training and the rest for testing and the accuracy is obtained with five-fold cross validation. The achieved results are acceptable for medium security applications and the segmentation process proved to be not necessary as it may remove important parts in the image background. More statistical features can be added in future work with the aim of improved accuracy. To compare the achieved results with the state-of the-Art models, this method is compared with the method used in [14] on the AMI database. In [14] The LBP is used and achieved an accuracy of 73 ± 1.88 which is lower than the achieved accuracy presented in this work which proves the efficiency of the proposed method using handcrafted features.

In the second scenario, deep learning is employed. First, end-to-end using three pretrained nets which are AlexNet, GoogleNet and ResNet50 is performed. Two ear datasets were investigated which are the AMI with 100 classes each having seven images and the IIT Delhi1 with 125 classes for raw and segmented images each having at least three images. To overcome the limited data size, data augmentation is used. Transfer learning is applied to adapt the final layers to the required task. The best achieved accuracy for the end-to-end experiments reached 99% with ResNet50 for the AMI database. The AlexNet Provided the best results for the IIT Delhi database for both raw and segmented images with accuracies 94.29% and 62.89% respectively.

End-to-end classification is considered the best option nowadays as the accuracy is calculated in one step with no pre-processing.

Features are then extracted from the chosen CNNs and are passed to a shallow classifier. Different classifiers were investigated but the only classifiers which succeeded to give superior results were the ensemble classifiers especially the subspace discriminant. PCA is then the applied to reduce the feature vector, which is again passed to the classifier. Different variance

Table 21 Percentage classification results for features of GoogleNet with enseble classifiers on IIT Delhi segmented database

# of Input Features	Ensemble bagged	Subspace discriminant	Subspace knn
1024	11.6	33.5	16.4
97 PCA with 95% var	8.7	33.3	15.6
140 PCA with 97% var	7.3	35.1	15.6
249 PCA with 99% var	7.7	27.6	16

Table 22 Percentage classification results for features ResNet50 with ensemble classifiers on IIT Delhi segmented database

# of Input Features	Ensemble bagged	Subspace discriminant	Subspace knn
1024	13	38.1	16.8
123 PCA with 95% var	11.6	50.9	15.2
178 PCA with 97% var	12.4	49.1	15.8
296 PCA with 99% var	8.9	26	16.8

levels are investigated which are 95%, 97% and 99%. For the AMI database, the best obtained results are for the full length of the feature vector of 2048 ResNet50 coefficients and reached 99.45%. Second best was 98.1% for AlexNet features with 4096 coefficients and third best was 97,3% with 97% variance with 271 coefficients. As for the IIT Delhi raw image database, the AlexNet features reduced with PCA with 95% variance gave the best result with 93.9% accuracy with 117 coefficients. Same accuracy was with 1024 google net coefficients and then 93.5% for Resnet50 coefficients reduced with PCA with 95% variance.

The worst obtained results was for the IIT Delhi segmented database with a maximum accuracy of 50.9% with ResNet50 coefficients reduced with PCA at 95% variance. Other results were lower than these values.

The achieved results for the segmented database confirm the idea that the segmentation process may remove important parts from the image which may degrade the accuracy results.

To compare the results with the state-of-the-art methods, we can find that the proposed method outperformed all methods using deep learning on the AMI dataset with a mean accuracy of 99% for end-to-end classification and 99.45% using Ensemble classifiers on extracted deep features. The achieved results are compared to the state-of-the-art models in [21] which used DCGAN + VGG16, and [3] which used Ensembes of VGG 13-16-19, and achieved an accuracy of 96% in the first and 97.5% in the latter, we find that the proposed model produced superior results with an accuracy of 99% for end-to-end classification and 99.45% with ensemble classifiers. This is not the case with the IIT Delhi database as the achieved accuracy was lower than the state-of-the-art methods as provided in Table 21. Some suggestions to improve the accuracy of the IIT Delhi database might be using other deepnets such as Dense and Dark nets, combining deep features with handcrafted features or combining deep features at different levels.

The performance of proposed method is compared with previous techniques provided in literature on the same used databases and the results are given in Table 24.

The comparative table shows that the proposed method produced superior results for the AMI database.

Table 23 t test results for the Alexnet features reduced with PCA and passed to ensemble classifiers for IIT Delhi database

t-score	SED	DOF	MD	CR	P value
858.96	0.1093	9	93.92	93.67–94.16	<0.05

Table 24 Comparison with previous methods

Summary of related work	Method	Classifier	AMI	IIT Delhi1
Amir Benzaou elal, 2014 [6]	BSIF descriptor	KNN	-	97.6
A. Kumar, C. Wu, 2011 [24]	Orthogonal log-Gabor filter Pair	KNN	-	96,27
A. Basit & M. Shoaib, 2014 [1]	Curvelet features	KNN		97.77
Ibrahim Omara et al., 2018 [18]	CNN features+DCA	Pairwise SVM		99.5
Yacine Khaldi1 · Amir Benzaoui2, 2020 [21]	DCGAN+VGG16		96	
Hammam Alshazly et al., 2019 [3]	Ensembles of VGG-13-16-19		97.5	
M. Hassaballaha et al., 2019 [14]	Local binary patterns	Chi-square dissimilarity measure	73.71±2.61	97.16±1.35
Ramar Ahila Priyadharshini et al., 2020 [33]	Six layer deep net		96.99%	97.36%
The proposed method	Curvelet features	Ensemble	86.5	
	Resnet50	End-to-end	99	93.57
	Alexnet	End-to end	94	94.29
	Resnet features	Ensemble	99.45	93.5 with PCA(95% var)

8 Conclusions

Two tracks for ear recognition are investigated in this paper. In the first scenario, the ear images are segmented and then statistical features are extracted from the discrete curvelet transform at different levels to form the feature vector which is then passed to the ensemble classifiers to obtain the recognition accuracy. Non-segmented ear images are also investigated. The classification accuracy for segmented ear images was higher than that of non-segmented images by about 3% which raised the question of the necessity of the segmentation process. In the second scenario, deep learning methods are employed in an end-to-end procedure and then features are passed to a shallow classifier in another procedure. The extracted features are reduced with PCA. This process was implemented on two databases with segmented and non-segmented ear images. The non-segmented images provided superior results over the segmented images for both methods especially with the subspace discriminant ensemble classifiers. The proposed method produced superior results for the AMI database.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Declarations

Conflict of interest The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Abdul B, Shoaib M (2014) A human ear recognition method using nonlinear curvelet feature subspace. Int J Comput Math 91(3):616–624
2. Alqaralleh E, Toygar O (2018) Ear recognition based on fusion of ear and tragus under different challenges. Int J Pattern Recognit Artif Intell
3. Alshazly H, Linse C, Barth E, Martinetz T (2019) Ensembles of deep learning models and transfer learning for ear recognition. Sensors
4. Barnabas Victor K, Bowyer, Sarkar S (2002) An evaluation of face and ear biometric. 16th International conference of Pattern Recognition, pp 429–432
5. Barnabas Victor K, Bowyer S, Sarkar (2002) An evaluation of face and ear biometrics. In: Proceedings of the International Conference on Pattern Recognition, vol 1. IEEE, pp 429–432
6. Benzaoui A, Hadid A, Boukrouche A (2014) Ear biometric recognition using local texture descriptors. J Electron Imaging
7. Burge M, Burger W (1998) Ear biometrics, biometrics: personal identification in networked society. In: Jain AK, Bolle R, Pankanti S (eds), pp 273–286
8. Bustard JD, Nixon MS (2010) Toward unconstrained ear recognition from two-dimensional images. IEEE Trans Syst Man Cybern Part A: Syst Hum 40(3)

9. Candès E, Donoho DL (1999) Curvelets—A surprisingly effective nonadaptive representation for objects with edges. In: Curve and Surface Fitting: Saint-Malo
10. Candes E, Demanet L, Donoho D, Ying L (2006) Fast discrete curvelet transforms. SIAM J Multiscale Model Simul
11. Dodge S, Mounsef J, Karam L (2018) Unconstrained ear recognition using deep neural networks. IET Biom 7:207–214
12. Eyiokur FI, Yaman D, Ekenel HK (2018) Domain adaptation for ear recognition using deep convolutional neural networks. IET Biom 7(3):199–206
13. Hansley EE, Segundo P, Sarkar S (2018) Employing fusion of learned and handcrafted features for unconstrained ear recognition. IET Biom 7(3):215–223
14. Hassaballaha M, Alshazly HA, Ali AA (2019) Ear recognition using local binary patterns: A comparative experimental study. Expert Syst Appl
15. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA
16. Hurley DJ, Nixon MS, Carter JN (2000) Automatic ear recognition by force field transformations. In: Proceedings of the Colloquium on Visual Biometrics, IET, pp 7–1
17. Hurley DJ, Nixon MS, Carter JN (2005) Ear biometrics by force field convergence. In: Proceedings of the Audio-and Video-Based Biometric Person Authentication. Springer, pp 386–394
18. Ibrahim Omara X, Wu H, Zhang Y, Du W, Zuo (2018) Learning pairwise SVM on hierarchical deep features for ear recognition. IET Biom 7(6):557–566
19. Ivo Alberink A, Ruifrok (2007) Performance of the Fear ID earprint identification system. Forensic Sci Int 166(2):145–154
20. Kacar U, Kirci M (2019) ScoreNet: deep cascade score level fusion for unconstrained ear recognition. IET Biom 8(2)
21. Khaldi Y, Benzaoui A (2020) A new framework for grayscale ear images recognition using generative adversarial networks under unconstrained conditions. Evol Syst
22. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst
23. Kumar A, Zhang D (2007) Ear authentication using log-gabor wavelets. In: Proceedings of the Symposium on Defense and Security, International Society for Optics and Photonics, p 65390A
24. Kumar A, Wu C (2011) Automated human identification using ear imaging. Pattern Recognit
25. Kumar K, Shrimankar DD (2018) F-DES: Fast and deep event summarization. IEEE Trans Multimed 20(2)
26. Kyong C, Barnabas V (2003) Comparison and combination of ear and face image in appearance-based biometrics. IEEE Trans Pattern Anal Mach Intell 25:1160–1165
27. Moreno B, SÁnchezA, Vélez JF (1999) On the use of outer ear images for personal identification in security applications. In: Proceedings of the International Carnahan Conference on Security Technology, IEEE, pp 469–476
28. Nosrati MS, Faez K, Faradji F (2007) Using 2D wavelet and principal component analysis for personal identification based on 2D ear structure. In: Proceedings of the International Conference on Intelligent and Advanced Systems. IEEE, pp 616–620
29. Olanrewaju L, Oyebiyi O, Misra S, Maskeliunas R, Damasevicius R (2020) Secure ear biometrics using circular kernel principal component analysis, Chebyshev transform hashing and Bose–Chaudhuri–Hocquenghem errorcorrecting codes. Signal Image Video Process
30. Pflug A, Busch C, Ross R (2014) 2D ear classification based on unsupervised clustering. In: Proceedings of the International Joint Conference on Biometrics. IEEE, pp 1–8
31. Pflug A, Paul PN, Busch C (2014) A comparative study on texture and surface descriptors for ear biometrics. In: Proceedings of the International Carnahan Conference on Security Technology. IEEE, pp 1–6
32. Polikar R (2006) Ensemble based systems in decision making. IEEE Circuits Syst Mag 6(3)
33. Priyadharshini RA, Arivazhagan S, Arun M (2020) A deep learning approach for person identification using ear biometrics. Appl Intell
34. Rahman A, Tasnim S (2014) Ensemble classifiers and their applications: a review. Int J Comput Trends Technol (IJCTT) 10(1)
35. Sharma S, Kumar K, Singh N (2020) Deep eigen space based ASL recognition system. IETE J Res
36. Sumana IJ, Islam MdM, Zhang D, Lu G (2008) Content based image retrieval using curvelet transform. IEEE 10th Workshop on Multimedia Signal Processing

37. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D et al. (2015) Going deeper with convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA
38. Wang Z-Q, Yan X-d (2011) Multi-scale feature extraction algorithm of ear image. In: Proceedings of the International Conference on Electric Information and Control Engineering. IEEE, pp 528–531
39. Yuan L, Mu Z-c, Zhang Y, Liu K (2006) Ear recognition using improved non-negative matrix factorization. In: Proceedings of the International Conference on Pattern Recognition, vol 4. IEEE, pp 501–504
40. Zhang Y, Mu Z, Yuan L, Yu C (2018) Ear verification under uncontrolled conditions with convolutional neural networks. IET Biom 7:185–198