Hindawi
Applied Computational Intelligence and Soft Computing
Volume 2022, Article ID 9692690, 17 pages
https://doi.org/10.1155/2022/9692690

Review Article

Ear Biometrics Using Deep Learning: A Survey

Aimee Booysens and Serestina Viriri

School of Mathematics, Statistics & Computer Science, University of KwaZulu-Natal, Durban, South Africa

Correspondence should be addressed to Serestina Viriri; viriris@ukzn.ac.za

Received 17 February 2022; Accepted 11 July 2022; Published 17 August 2022

Academic Editor: Agostino Forestiero

This paper explores ear biometrics using a mixture of feature extraction techniques and classifies this feature vector using deep learning with convolutional neural network. This exploration of ear biometrics uses images from 2D facial profiles and facial images. The investigated feature techniques are Zernike Moments, local binary pattern, Gabor filter, and Haralick texture moments. The normalised feature vector is used to examine whether deep learning using convolutional neural network is better at identifying the ear than other commonly used machine learning techniques. The widely used machine learning techniques that were used to compare them are decision tree, naïve Bayes, K-nearest neighbors (KNN), and support vector machine (SVM). This paper proved that using a bag of feature techniques and the classification technique of deep learning using convolutional neural network was better than standard machine learning techniques. The result achieved by the deep learning using convolutional neural network was 92.00% average ear identification rate for both left and right ears.

1. Introduction

The ear begins to develop on a fetus amid the fifth and seventh weeks of pregnancy [1]. At this stage of the pregnancy, the face acquires a more distinguishable shape as the mouth, nostrils, and ears begin to form. There is still no exact timeline at which the outer ear is created during pregnancy, but it is accepted that a cluster of embryonic cells connect to establish the ear. These are called auricular hillocks, which begin to grow in the lower portion of the neck. The auricular hillocks broaden and intertwine within the seventh week to deliver the ear's shape. Within the ninth week, the hillocks move to the ear canal and are more noticeable as the ear [1]. The external anatomy of the ear can be seen in Figure 1. The growth of the ear in the first four months after birth is linear. The ear is then stretched in development between the ages of four months and eight years. After this, the ear size and shape are constant until the age of seventy, when they increase in size again.

Biometrics is the recognition of a human using their biometric characteristics, which may be physiological or behavioural. The physiological biometric features are the DNA, face, ear, facial, iris, fingerprint, hand geometry, hand vein, and palm print, with the behavioural biometrics being signatures, gait pattern, and keystrokes. Voice is considered as a combination of biometric and physiological. Numerous systems have been developed to distinguish biometric traits, which have been used in numerous applications such as forensic investigations and security systems. With the present worldwide pandemic, facial identification has failed due to users' wearing masks. However, the human ear has proven more suitable as it is visible. In Table 1, the characteristics that were looked at were the performance of the biometric if it is distinctive, permanence, ability to be collected, and acceptability.

In the different physiological biometric qualities, the ear has received much consideration of late as it tends to be said that it is a solid biometric for human acknowledgement [2]. The ear biometric framework is dependable as it does not change, it is of uniform tone, and its position is fixed at the centre of the face's side. The size of an individual's ear is more critical than a unique finger impression and makes it simpler to capture an image of the subject without necessarily needing to gain information from the subject [2]. There are numerous difficulties in correctly gauging the details of the ear. These are concealment of the ear by clothes,

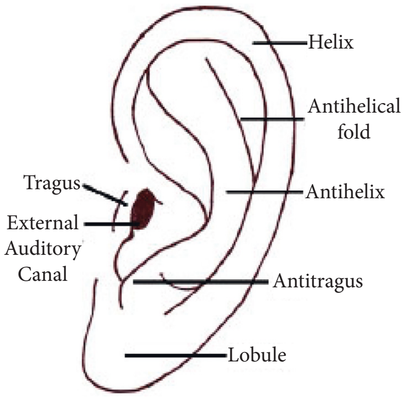


Figure 1: Diagram of the outer ear.

Table 1: Summary of biometric characteristics.

Biometric identifier	Biometric type	Distinctiveness	Permanence	Collectability	Performance	Acceptability
DNA	Physiological	High	High	Low	High	Low
Ear	Physiological	Medium	High	Medium	Medium	High
Face	Physiological	Low	Medium	High	Low	High
Facial	Physiological	High	Low	High	Medium	High
Fingerprint	Physiological	High	High	Medium	High	Medium
Gait	Behavioural	Low	Low	High	Low	High
Hand geometry	Physiological	Medium	Medium	High	Medium	Medium
Hand vein	Physiological	Medium	Medium	Medium	Medium	Medium
Iris	Physiological	High	High	Medium	High	Low
Keystroke	Behavioural	Low	Low	Medium	Low	Medium
Odor	Physiological	High	High	Low	Low	Medium
Palm print	Physiological	High	High	Medium	High	Medium
Retina	Physiological	High	Medium	Low	High	Low
Signature	Behavioural	Low	Low	High	Low	High
Voice	Combination of physiological and behavioural	Low	Low	Medium	Low	High

hair, ear ornaments, and jewellery. Another inference could be the different angle at which the image was taken, concealing essential characteristics of the ear's anatomy. These difficulties made ear recognition a secondary role in identification systems and techniques commonly used for identification and verification.

This paper's contributions are summarised below.

(1) A survey has been conducted with different deep learning architectures

(2) A study of the present ear bench-mark databases and their suitability for ear identification

(3) Different algorithms used for ear identification were outlined, highlighting the weaknesses and strengths

(4) A review of the present deep learning algorithms used for ear identification

The remainder of this work is organised as follows: Section 2 presents the foundation data on deep learning; Section 3 presents the vast majority of the ear information bases that are accessible for research; Section 4 presents a study of ear recognition calculations; the different profound learning strategies used to identify the ear are introduced in Section 5; and Section 6 presents the conclusion.

2. Review of Deep Learning

Deep learning is an AI model that utilises numerous layers to progressively understand the data. This paper will discuss the structures and contemporary strategies for deep learning designs in AI models that find the correct representation for the inputted information.

2.1. Neural Network (NN). A neural network (NN) is a type of machine learning algorithm that learns representations from data [3, 4]. A neutron may connect the processing unit from the directly linked network. Whenever there is a link, it has a weight that will be adjusted to assist the training process. The feed-forward neural network is when each neuron may be a function $f(x\colon \theta)$ which maps to an input, then to an output. The network learns the values of the parameters $\theta = w, b$, where w is a weight vector and b a scalar. This is often performed through a backpropagation algorithm, as shown in the following equation:

$$F(x\colon\ \theta) = \sigma(w.x + b). \tag{1}$$

The first layer within the network is the input layer, and therefore, the last layer is the output layer. The middle layers within the algorithm are referred to as the hidden layers. When there are many hidden layers, this is often mentioned as a deep neural network; this is depicted in Figure 2.

2.2. Convolutional Neural Network (CNN). A convolutional neural network (CNN) is an NN that joins two or more layers together to produce one composite layer. The convolutional layer is able to learn features from the input data. By stacking many convolutional layers, the network is able to learn a hierarchy of increasingly complex features [3]. A pooling layer is usually added between successive convolutional layers to reinforce essential elements. In doing the CNN, it reduces the number of parameters that are passed to the lower layers. This is depicted in Figure 3.

2.3. Building Block for Convolutional Neural Networks

2.3.1. Convolutional Layer. This layer is a set of learnable filters or kernels used to slide over the entire input volume, performing a dot product between entries of the filter and the input layer [5]. The convolutional operation first extracts patches from its information in a sliding window fashion and then applies the same linear transformation to all the areas. The output of the convolutional operation is referred to as a feature map. The network will learn filters and then recognise the visual patterns that are in the input data. This is often shown asx_{ij}^{l}

$$x_{ij}^{l} = \sum_{a=0}^{m-1}\sum_{b=0}^{m-1} \omega_{ab} y_{(i+a)(j+b)}^{l-1}, \tag{2}$$

where x_{ij}^{l} is the computation of the input and is the sum of the contributions from the previous layer cells.

2.3.2. Pooling Layer. A pooling layer usually follows a single or multiple convolutional layers and is used to reduce the feature mapped dimensions keeping the essential elements [3]. A pooling layer is applied to a rectangular neighbourhood using a sliding window operation. Other pooling operations are maximum, depicted in Figure 4, average depicted in Figure 5, and weighted global pooling.

2.3.3. Nonlinearity Layer. The nonlinearity layer involves three steps. In step one, the layer performs the convolutional operation on the input feature map and produces a linear activation [3]. The second step would be to do the nonlinear transformation, and lastly, the pooling layer is used to modify the output. Nonlinear transformation can be carried out using activation functions; this gives the network the ability to learn a nontrivial representation, making the network resilient to slight modifications or noise in the input data and improving the computational efficiency. This is often shown as$lY_i^{(l-1)}l-1$

$$Y_i^{(l)} = f\left(Y_i^{(l-1)}\right), \tag{3}$$

where l is the nonlinearity layer and the volume $Y_i^{(l-1)}$ is from the convolutional layer $l-1$.

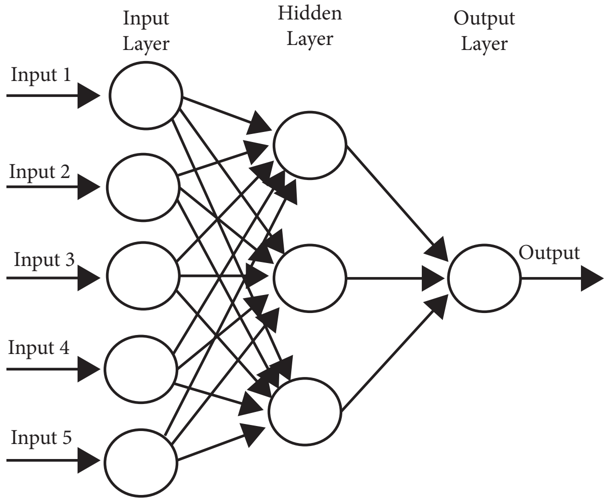


Figure 2: Diagram of neural networks.

2.3.4. Fully Connected Layer. The fully connected layer is used as a feature extractor. The features produced are then passed to the fully connected layers for classification. Each unit in the fully connected layer is connected to all the units in the previous layers. The last layer is usually a classifier that produces a probability map over the different classes. All the features are converted into one-dimensional feature vectors before passing into the fully connected layer. The reason that this is carried out is that spatial information in the image data is lost, has a high computational cost, and can only work with images that are of the same size [6]. This is often shown as

$$\begin{aligned} y_i^{(l)} &= f\left(z_i^{(l)}\right)\text{with } z_i^{(l)} \\ &= \sum_{j=1}^{m_1^{l-1}} w_{i,j}^{(l)} y_i^{(l-1)}. \end{aligned} \tag{4}$$

2.3.5. Optimisation. The performance of the deep CNN can be improved by training the network on a large data set. Training involves looking for the parameter of the model that reduces the cost function [3]. Gradient descent, shown in equation (5), is a widely used method for updating the network parameters through the backpropagation algorithm. The optimisation can be carried out at any stage in the process.

$$\Theta = \Theta - \alpha \cdot \nabla J(\Theta). \tag{5}$$

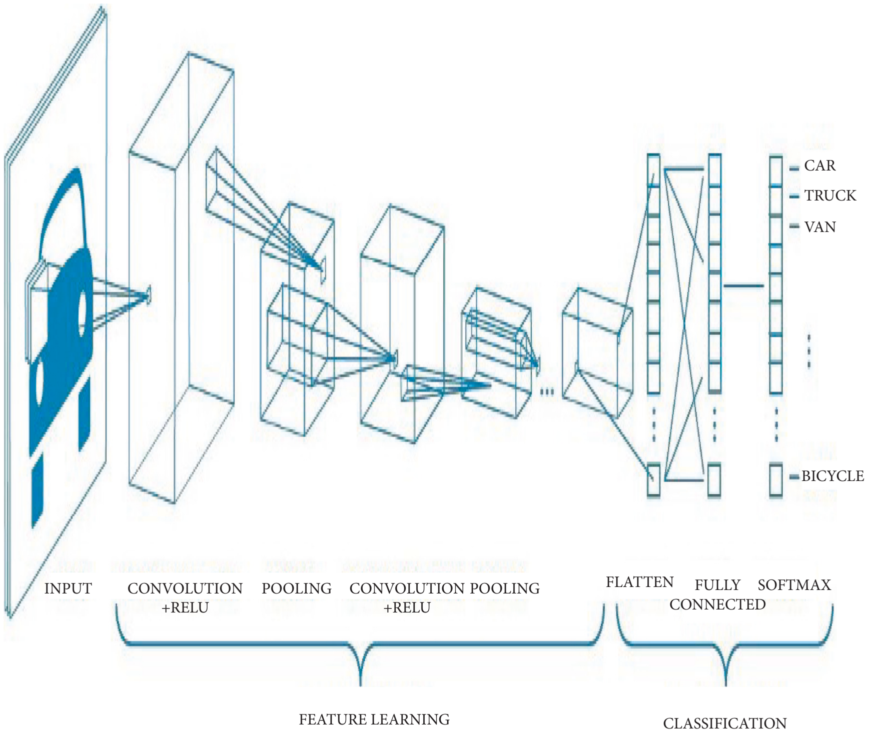
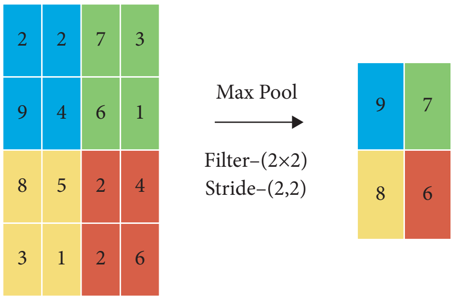


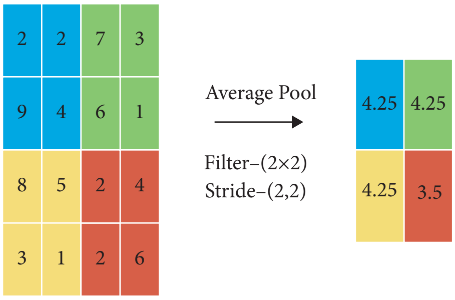
Figure 3: Diagram of convolutional neural networks.



2	2	7	3
9	4	6	1
8	5	2	4
3	1	2	6

9	7
8	6

Figure 4: Diagram of maximum pooling layer.



2	2	7	3
9	4	6	1
8	5	2	4
3	1	2	6

4.25	4.25
4.25	3.5

Figure 5: Diagram of average pooling layer.

2.3.6. Loss Function. Loss function is used in machine learning to evaluate how the specific algorithm model obtains data. The main goal of training an NN is to make sure that the loss is low. When the output is far from the actual value, the loss will be high and low when the prediction is close to the actual value [3]. The loss function used is mean-squared error, which is calculated by taking the mean of squared differences between actual and predicted values, and the binary cross entropy takes the output node to classify the data into two classes which are passed through a sigmoid function with an output of 0 or 1.

2.3.7. Parameter Initialisation. Parameter initialisation is a deep learning optimisation algorithm that is iterative and requires the user to state a starting point for the algorithm. The point at which the user chooses influences how fast learning can converge [3].

2.3.8. Hyperparameter Tuning. Hyperparameter tuning is the parameter that the user supplies to control the algorithm's behaviour before training starts, and this can be the learning rate, batch size, or image size [3].

2.3.9. Regularisation. Regularisation is a technique for improving the performance of machine learning algorithms on unseen data [3]. Regularisation is carried out to reduce the overfitting of the training set, and this happens when the gap between the training and test error is too large.

2.4. Deep Convolutional Neural Network Architectures

2.4.1. Single Pathway. A single pathway may be a primary network that resembles a feed-forward deep neural network [7]. Using one path, the data moves from the input layer to the classification layer. Kleesiek et al. [8] proposed a 3D single-path CNN that has fully connected convolutional layers: the classification layer, which allows the network to classify multiple 3D pixels on just one occasion.

2.4.2. Cascaded Architecture. In the cascaded architecture, the output of the CNN is concatenated with another [9]. There are many variations with this architecture within the literature, but the input cascade is prominent. In this architecture, the output of the CNN becomes a direct input of another CNN. The input cascade is employed to concatenate the contextual information to the second CNN as additional image channels. Cascaded architecture is an improvement to the only pathway that performs multiscale label prediction separately. There are many other cascaded architectures: local pathway concatenation and hierarchical segmentation.

2.4.3. UNET. UNET improves a convolutional network that resembles an encoder and decoder network designed to do biomedical image segmentation [10]. The network consists of a contracting path and an expansive path, which provides it with the u-shaped architecture. The contracting path consists of the repeated application of two convolutional layers, followed by a rectified linear measure and a top pooling layer that goes along the trail to scale back the spatial information while feature information is increased. The expansive path consists of upsampling operations combined with high-resolution features from the contraction path through skip connections.

2.4.4. AlexNet Architecture. AlexNet architecture is an easy but powerful CNN architecture consisting of convolutional and pooling layers [11]. These layers are fully connected at the highest point, and the benefits of the AlexNet include the size with which it uses the GPU for training and performing the task. This architecture remains a starting point in applying deep neural networks, specifically for computer vision and speech recognition.

2.4.5. Visual Geometry Group Architecture. Visual geometry group architecture is a network created by Visual Graphics Group researchers at Oxford University [12]. It is characterised by a pyramidal shape because it comprises a group of convolutional layers followed by pooling layers; these pooling layers make the layers narrower in shape. The benefits include keeping a good architecture used for benchmarking for any task. The pretrained networks of the VGG are also primarily used for different applications but require numerous computational resources and are slow to coach, above all when training the dataset from scratch.

2.4.6. GoogLeNet Architecture. The GoogLeNet architecture is referred to as the inception network and was created by Google researchers [13]. It is made from twenty-two layers with two options that these layers can either convolute or pool the input. The architecture contains many beginning modules stacked over each other, allowing joint and parallel training, which helps with faster convergence. The benefits are that there is speedier training, which reduces the size. It , however, possesses an Xception network, which could increase the point for the divergence of the beginning module.

2.4.7. Residual Network (ResNet) Architecture. The residual network (ResNet) architecture is a 152-layer deep CNN architecture of the residual blocks. This is more profound than that of the AlexNet and VGG architectures as it is less computationally complex than these networks. It is referred to as a residual network [14], which is made up of numerous succeeding residual modules that are the essential building blocks of the architecture. These modules are stacked to produce an end-to-end network. The advantage of this architecture is that performance is improved due to its many residual layers and it is used for network training.

2.4.8. ResNeXt Architecture. ResNeXt architecture is the present state-of-the-art technique for visual perception, which is a hybridisation between inception and ResNeXt architectures [15]. ResNeXt is referred to as the aggregated residual transform network, but it is an improvement over the inception network. It splits the concept and transforms and merges in a commanding but easy way by bringing in cardinality. It uses residual learning, which will enhance the joining of the deep and wide networks. ResNeXt uses many transformations within a split, transform, and merge blocks; and the transformations in cardinality define these. ResNeXt used a mixture of VGG topology and GoogLeNet architecture to correct the spatial resolution using 3×3 filters within the split, transform, and merge blocks. The increase in cardinality improves the performance and produces a different and improved architecture.

2.4.9. Advance Inception Network. The advance inception network includes Inception-V3, Inception-V4, and Inception-ResNet. This is often an improved version of Inception-V1, Inception-V2, and GoogLeNet [16]. Inception-V3 reduces the computational cost of deep networks but does not affect generalisation. Szegedy et al. [17] replaced large-sized filters (5×5 and 7×7) with small and unequal filters (1×7 and 1×5) and used 1×1 convolution as a blockage before the vast filters. Inception-ResNet combines the strength of the residual learning and starting block.

2.4.10. DenseNet Architecture. The DenseNet architecture [16] is similar to ResNet but was created to fix the vanishing gradient problem. DenseNet utilises cross-layer connectivity by connecting each preceding layer to the next layer in a feed-forward manner. This was carried out to fix the ResNet by preserving identity transformations, which increased complexity. As it uses solid blocks, it allows to feature maps of all previous layers to be used as the inputs into the subsequent layers.

2.4.11. SqueezeNet Architecture. Hu et al. [18] proposed an auxiliary block for the choice to feature maps for object discrimination. The new block named SE-block overpowers the smaller feature maps and stimulates the category feature maps. It was created to be added into any CNN architecture before the convolution layer. It has two primary operations: squeeze and convolution. The convolution kernel captures local information but ignores features' contextual relations, while the squeeze operation captures global information of the feature maps. The network generates a feature map that is a more robust architecture and is helpful when there is low bandwidth.

2.4.12. Xception Architecture. Xception architecture is referred to as risky inception architecture that overdoes depthwise separable convolution [19]. The first inception block is modified by making it more complete and substituting different spatial dimensions (1 × 1, 5 × 5, and 3 × 3) with one dimension (3 × 3) followed by a 1 × 1 convolution to achieve computational complexity. It makes the network computationally efficient by uncoupling spatial and feature map channels.

2.4.13. Deep Reinforcement Learning. Deep reinforcement learning [20] may be a system trained entirely from scratch, ranging from random behaviour to an accurate knowledge domain from experience. It is a mixture of reinforcement and deep learning using fewer computation resources and data. The algorithm can learn from its environment and apply it to any sequential decision-making problems, including image analysis.

2.4.14. Fully Convolutional Network. A fully convolutional network [21] is a set of convolutional and pooling layers. Bi et al. [22] developed a multistage fully convolutional network with the parallel integration method for segmentation.

2.4.15. Deep Residual Network. Deep residual network [23] may be a particular sort of artificial neural network that builds on a pyramidal structure by utilising skip connections that skip some convolutional layers. It is composed mainly of multiple convolutional layers.

2.4.16. Convolutional and Deconvolutional Neural Networks. This architecture is formed from two significant parts: convolutional and deconvolutional networks [24]. Deconvolutional networks are CNNs that operate during a reversed process, and networks extract discriminated features. The deconvolutional layers are applied for smothering the segmentation maps to get the ultimate high-resolution output.

2.4.17. Residual Attention Neural. Zhou et al. [25] designed residual attention neural that improves CNNs feature representation by incorporating attention modules into CNN and forms a network capable of learning object-aware features. It employs a feed-forward CNN that stacks residual blocks with an attention module. It combines two different learning strategies into the eye module that permits fast feed-forward processing and top-down attention feedback during a single feed-forward process to supply dense features that infer each pixel. The bottom-up feed-forward structure produces low-resolution feature maps with reliable semantic information. The top-down learning strategy globally optimises the network such that it gradually outputs the maps to input during the training process. Table 2 shows a summary of the deep convolutional neural network architecture used for ear identification.

3. Overview of the Ear Dataset

Many factors can affect an ear detection system's performance. The ear images' datasets are easier to use than others. The more ear datasets are for researchers to use, the more this field can evolve and grow. It is always good to use high-quality images in research associated with soft biometrics. A brief description of a number of the available ear databases is highlighted in Table 3 and examples of images are shown in Figures 6 and 7.

3.1. Mathematical Analysis of Images (AMI) Ear Database. The AMI ear database was collected at the University of Las Palmas. The database comprises 700 ear images of 100 distinct Caucasian adult males and females between 19 and 65 years of age. All images within the database were taken under equivalent illumination and with a glued camera position. Both the left- and right-hand sides of the ears were captured. The pictures obtained are cropped to form the ear area, covering almost half of the image. The pose of the themes varies in yaw and surveying in pitch angles, and datasets are often found publicly.

3.2. The Indian Institute of Technology (IIT) Delhi Ear Database. The IIT database [26] was collected by the Indian Institute of Technology Delhi in New Delhi between October 2006 and June 2007. The database is formed from 421 images of 121 distinct adults of both males and females. All images were taken inside the environment, with no significant occlusions present, and only the right-hand side of the ear was captured. The pictures obtained in the dataset were both raw and normalised. The normalised images were in greyscale with a size of 272 × 204 pixels.

3.3. The University of Beira Ear (UBEAR) Database. The University of Beira presented the UBEAR database [27]. The database comprises 4429 images of 126 subjects, and these were of both males and females. The images were taken under varying lighting conditions and angles, and partial occlusions were present. These images are of the ear, both the left- and right-hand side ear images were provided.

3.4. The Annotated Web Ear (AWE) Database. The AWE ear database [28] was a set of public figures from web images. The database was formed from 1000 images of 100

Table 2: Summary of the deep convolutional neural network architecture used for ear identification.

Deep convolutional neural network architecture	Summary of deep convolutional neural network used in ear identification	Accuracy (%)
AlexNet [11]	AlexNet is seen as a deep convolutional neural network architecture and applied to numerous ear recognition systems	53.6
DenseNet [16]	DenseNet connects each layer in the CNN to another and applied to ear image datasets, yielding positive results	62.0
ResNet [14]	ResNet is a class of extremely deep CNN architecture that addresses vanishing gradient by using skip connections that prevent information loss as the network goes deeper. As ResNet addressed the vanishing gradient issue, it has been applied to numerous ear image datasets yielding positive results	15.0
ResNeXt [15]	ResNeXt is a modularised CNN architecture, which has been applied to ear image datasets yielding positive results	95.8
Visual geometry group [12]	The visual geometry group is a very deep CNN and is one of the top performers. The VGG is used in recognition systems and has been applied to unconstrained ear image datasets, yielding positive results.	83.0

different subjects, whose sizes varied and were tightly cropped. Both the left- and right-hand sides of the ears were taken.

3.5. EarVN1.0. The EarVN1.0 database [29] comprises 28412 images of 164 Asian male and female subjects, and left- and right-hand sides of the ear were captured. It was collected during 2018 and is formed from unconstrained conditions, including camera systems and lighting conditions. The pictures are cropped from facial images to obtain the ears, and the pictures have significant variations in pose, scale, and illumination.

3.6. The Western Pomeranian University of Technology Ear (WPUTE) Database. The Western Pomeranian University of Technology Ear (WPUTE) database [32] was obtained in the year 2010 to gauge the ear recognition performance for images obtained in the wild. The database contains 2071 ear images belonging to 501 subjects. The images were of various sizes and held both the left- and right-hand sides of the ear and were taken under different indoor lighting conditions and rotations. There were some occlusions included in the database. These were the headset, earrings, and hearing aids.

3.7. The Unconstrained Ear Recognition Challenge (UERC). The Unconstrained Ear Recognition Challenge (UERC) database [14] was obtained in 2017, then extended in 2019, and is a mix of two databases that currently exist and a newly created one. The database contains 3706 subjects with 11804 ear images, and the database ears have both right- and left-hand side images.

3.8. In the Wild Ear (ITWE) Database. The In the Wild Ear (ITWE) database [33] was created for recognition evaluation and has 2058 total images, including 231 male and female subjects. A boundary box obtained these images of the ear. The coordinates of those boundary boxes were released with the gathering. The pictures contained cluttered backgrounds and were of variable size and determination. The database includes both the left- and right-hand sides of the ear, but no differentiation was given about the ears.

3.9. The University of Science and Technology, Beijing (USTB) Ear Database. The University of Science and Technology Beijing (USTB) Ear Database [30] contained cropped ear and head profile images of male and female subjects split into four sets. Dataset one includes 60 subjects and has 180 images of right-close-up ears during 2002. These images were taken under different lighting, experiencing some shearing and rotation. Dataset two contains 77 subjects and has 308 images of the right-hand side ear, approximately 2 m away from the ear, and the images were taken in 2004. These images were taken under different lighting conditions. Dataset three contains 103 subjects and has 1600 images. These images were taken during the year 2004. The images are on the proper and left rotation, and therefore, the images are of the dimensions 768 × 576. The dataset contains 25500 images of 500 subjects; these were obtained from 2007 to 2008; the subject was in the centre of the camera circle. The images were taken when the subject looked upwards, downwards, and at eye level. The images in this dataset contained different yaw and pitch poses. The databases are available on request and accessible for research.

3.10. The Carreira-Perpinan (CP) Ear Database. The Carreira-Perpinan (CP) [34] ear database is an early dataset of the ear utilised for ear recognition systems. It was created in 1995 and contained 102 images with 17 subjects. The images were captured in a controlled environment, and therefore, the images include variability in minor pose variation.

3.11. The Indian Institute of Technology, Kanpur (IITK) Ear Database. The Indian Institute of Technology Kanpur (IITK) is an ear database [35] that the Institute of Technology of Kanpur compiled. The database is split into three sets, the first set consists of 190 male and female subjects of profile images. The total number of images was 801. The second dataset also contained 801 total of 89 subjects, and

Table 3: Summary of datasets.

	Database	Year	Number of subjects	Number of images	Left ear count	Right ear count	Total ears	Image size	Country	Side
1	Institute of Technology Delhi Ear Database (IIT Delhi-I) [26]	2007	121	471		471	471	272 × 204	India	Right
	Institute of Technology Delhi Ear Database (IIT Delhi-II) [26]	NA	221	793		793	793	272 × 204	India	Right
2	The University of Science & Technology Beijing (USTB Ear I) [30]	2002	60	185		185	185	Varied	China	Right
	The University of Science & Technology Beijing (USTB Ear II) [30]	2004	77	308		308	308	Varied	China	Right
3	The Annotated Web Ears (AWE) database [28]	2016	100	1000	500	500	1000	Varied	Slovenia	Both
	The Annotated Web Ears database extended (AWE extend) [28]	2017	346	4104	2052	2052	4104	Varied	Slovenia	Both
4	Mathematical Analysis of Images Ear database (AMI) [31]	NA	106	700	420	280	700	492 × 702	Spain	Both
5	The West Pomeranian University of Technology Ear (WPUTE) database [32]	2010	501	2071	829	1242	2071	Varied	Poland	Both
6	Unconstrained Ear Recognition Challenge (UERC) database [14]	2017	3706	11804	5902	5902	11804	Varied	Slovenia	Both
7	EarVN1.0 [29]	2018	164	28412	14206	14206	28412	Varied and low-resolution	Vietnam	Both
8	The In-the Wild Ear (ITWE) database [33]	2015	55	605	424	181	605	Varied	Slovenia	Both
9	The Carreira-Perpinan (CP) [34]	1995	17	102	102		102	Varied	NA	Left
10	The University of Beira Ear (UBEAR) database [27]	2011	126	4430	2215	2215	4430	1280 × 960	Mozambique	Both
11	Indian Institute of Technology Kanpur (IITK) [35]	2011	801	190	95	95	190	Varied	India	Both
12	The forensic ear identification database (FEARID) [36]	2005	1229	1229	615	614	1229	Varied	United Kingdom, Italy and Netherlands	Both
13	University of Notre Dame (UND) [37]	2006	3480	952	952		952	Varied	France	Left
14	The Face Recognition Technology database (FERET) [38]	2010	9427	4745	3796	949	4745	Varied	Spain	Both
15	The Pose, Illumination, and Expression (PIE) [39]	2002	40000	68	34	34	68	Varied	USA	Both
16	The XM2VTS Ear Database [40]	NA	2360	295	89	206	295	720 × 576	UK	Both
17	The West Virginia University (WVU) [41]	2006	460	402	402		402	Varied	USA	Left

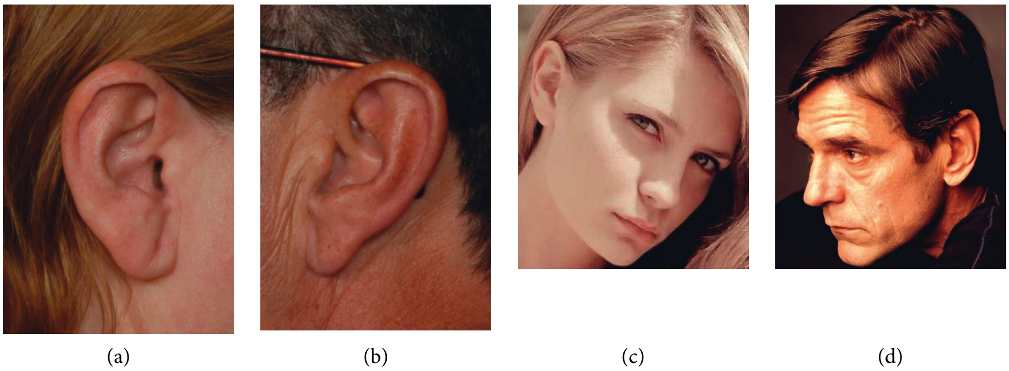


Figure 6: Examples of original ear images. (a) Example of a 2D profile image of a female. (b) Example of a 2D profile image of a male. (c) Example of a facial image of a female. (d) Example of a facial image of a male.

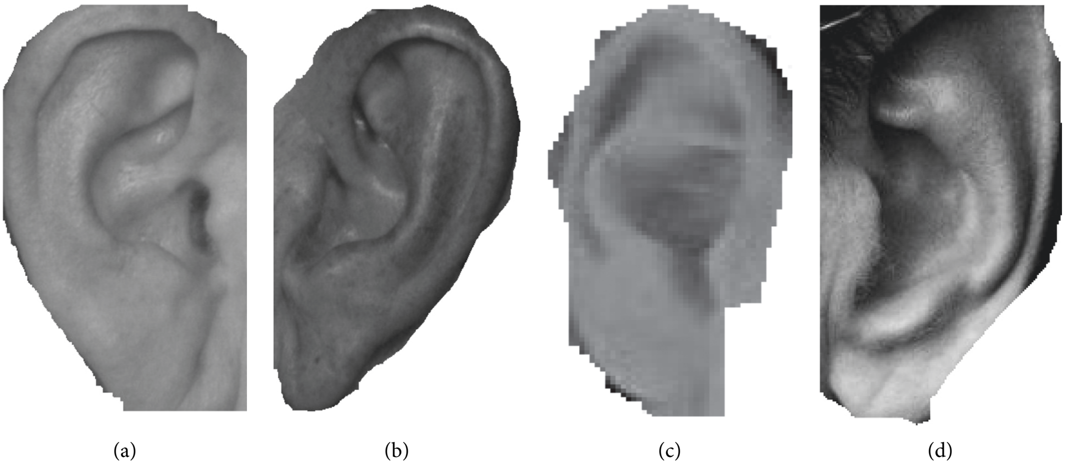


Figure 7: Examples of the extracted ear images. (a) Example ear extracted from 2D profile image of a female. (b) Example ear extracted from 2D profile image of a male. (c) Example ear extracted from facial image of a female. (d) Example ear extracted from facial image of a male.

these images had variations in pitch angle. The third dataset contains 1070 images of an equivalent of 89 subjects, but with a variation in yaw and angle.

3.12. The Forensic Ear Identification Database (FEARID). The Forensic Ear Identification Database (FEARID)[36] is different from other databases as it contains the ear prints. These contain no occlusions, variable angles, or illumination. Though there is no mention of any variables, other influences like the force the ear was pressed against the scanner and the scanner's cleanliness need to be considered. This database comprised 7364 images of 1229 subjects. This database was used for forensic application and not for biometric use.

3.13. The University of Notre Dame (UND) Database. The University of Notre Dame (UND) database contains [37] many subsets of 2D and 3D ear images. These images were appropriated for a period from 2003 to 2005. The database contains 3480 3D images from 952 male and female subjects and 464 2D images from 114 male and female subjects. These images were taken in different lighting conditions, yaw, pitch poses, and angles. The images are only of the left-hand side ear.

3.14. The Face Recognition Technology (FERET) Database. The Face Recognition Technology (FERET) database [38] is a sizeable facial image database and was obtained between the years 1995 and 1996. It contains 1564 subjects and has a total of 14126 images. These images were collected for face recognition and were of the left- and right-hand profile images, which made them perfect for 2D ear recognition.

3.15. The Pose, Illumination, and Expression (PIE). Carnegie Mellon University obtained the Pose, Illumination, and Expression database [39], which contains 40000 images and 68 subjects. The images are of the facial profile and have different poses, illuminations, and expressions.

3.16. The XM2VTS Ear Database. The XM2VTS ear database [40] is frontal and profiles face images from the University of Surrey; the database contains 295 subjects and 2360 images captured during controlled conditions. These images were a set of cropped images of 720 × 576 size and were from video data.

3.17. The West Virginia University (WVU) Ear Database. The West Virginia University (WVU) Ear database [41] is a video database and is formed from 137 subjects. The system was an advanced capturing procedure that allowed them to capture the ear at different angles; these images included earrings and eyeglasses.

3.18. Summary. UBEAR, EarVN1.0, IIT, ITWE, and AWE databases are best suited for the ear identification due to their large data size. However, it shows that EarVN1.0 has the foremost prominent usage during age estimation using CNN techniques. It is an appropriate dataset where the ear images are taken in a controlled environment, while ITWE is compatible for classifying the ears in an uncontrolled environment, and examples of the extracted ears are shown in Figures 6 and 7.

4. Description of Ear Algorithms

This section presents different algorithms and techniques used for ear identification. It presents a description of these algorithms and suggests the most effective approach. A brief description of ear algorithms is highlighted in Table 4.

Ansari and Gupta [42] used outer helix curves of the ears as they moved parallel to at least one feature spot in the ear image. Helix curves were obtained using the Canny edge detector to remove the ear from the entire image. The obtained sides are then separated into a convex or concave edge, allowing the system to determine the helix edges. This technique was run on 700 side-ear images and had an accuracy of roughly 93%.

Abdel-Mottaleb and Zhou [43] segmented the ear from a facial profile image using supported template matching, where they modelled the ear by its external curve. Yuizono et al. [44] also used a template matching technique for detection, in which they used both hierarchical 2D images. In 3D ear detection, Chen and Bhanu [45] used a model-based (template matching) technique for ear detection. An averaged histogram of the shape index represents the model template. The detection is a four-step process: edge detection and threshold, image dilation, connected component labelling, and template matching. A test set of 30 subjects from the UCR database achieved a 91.5% detection rate with a 2.52% warming rate. Later, Chen and Bhanu [45] developed another shape-model-based technique for locating human ears inside face range images, where the ear shape model is represented by a group of discrete 3D vertices like the helix and antihelix parts. They started by locating the sting segments and grouping them into different clusters that are potential ear candidates. Arbab-Zavar and Nixon [46] developed an ear recognition system based on the ear's elliptical shape, employing a Hough transformation (HT). They achieved a 100% detection rate using the XM2VTS face profile database, consisting of 252 images from 63 subjects, and 91% using the UND, collection F, database.

Burge and Burger [47] have proposed a way to do ear recognition using geometric information about the ear. The ear has been represented by employing a neighbourhood graph obtained from a Voronoi diagram of the ear edge segments, whereas template comparison has been performed using subgraph matching. Choras [48] has used the ear's geometric properties to propose an ear recognition technique during which feature extraction is administered in two steps. In the initial step, global features are extracted. The second step extracts local features while matching local features. In another geometry-based technique proposed by Shailaja and Gupta [49], an ear is represented by two sets of features, global and native, obtained using outer and internal ear edges, respectively. Two ears during this technique are declared similar if they are matched to the feature sets. The method proposed has treated the ear as a planar surface and has created a homograph transform using SIFT feature points to register ears accurately. It has achieved robust results in background clutter, viewing angle, and occlusion. Cummings et al. [50] used the image ray transformation, based upon an analogy to light rays, to detect an image's ears. This transformation can highlight tubular structures like the helix of the ear and spectacle frames. By exploiting the elliptical shape of the helix, this method segmented the ear into regions and achieved a detection rate of 99.6% using the XM2VTS database.

Chen and Bhanu [45] fused complexion from colour images and edges from a range of images to perform ear detection. The images observed that the sting magnitude is more prominent around the helix and, therefore, the antihelix parts. They clustered the resulting edge segments and deleted the short irrelevant edges. Using the UCR database, they reported an accurate detection rate of 99.3% (896 out of 902). The UND databases (collections *F* and a subset of *G*) reported an accurate detection rate of 87.71% (614 out of 700). Hajsaid et al. [51] addressed the matter of an automated ear segmentation scheme by employing morphological operators. They used low computational cost appearance-based features for segmentation and a learning-based Bayesian classifier to determine whether the segmentation's output was incorrect. They achieved a 90% accuracy on 3750 facial images with 376 subjects within the WVU database.

Prakash and Gupta [52] used complexion and template-based techniques for automatic ear detection during a side profile face image. The technique first separates skin regions from nonskin regions and then searches for the ear within the skin regions employing a template matching approach. Finally, the ear region is validated using a moment-based shape descriptor. Experimentation on an assembled database of 150 side-profile face images yielded an accuracy of 94% . Basrur et al. [53] introduced the notion of "jet space similarity" for ear detection, which denotes the similarity between Gabor jets and reconstructed jets obtained via principal component analysis (PCA). They used the XM2VTS database for evaluation; however, they did not report their algorithm's accuracy.

Rahman et al. [54] used a cascaded AdaBoost technique, supported by Haar features for ear detection. This system is

Table 4: Summary of the ear algorithms.

Author	Algorithms used	Accuracy (%)	Summary
Ansari and Gupta [42]	Canny edge detector	93	Uses outer helix curves of the ears with Canny edge detector, and this only obtains the edges of the ear and is only used to determine the helix
Abdel-Mottaleb and Zhou [43]	Template matching	91.5	They used a segmented ear obtained from a facial profile and only modelled the ear's external curve
Arbab-Zavar and Nixon [46]	Hough transform	91	They only looked at the ear's elliptical shape, and they used a small sample of profile ears
Burge and Burger [47]	Geometric information	94	They did ear recognition using geometric information of the ear and used neighbourhood graphs obtained from a Voronoi diagram of the ear edge segments
Cummings et al. [50]	Image ray transform	99.6	Used ray transformation to detect an image of the ear and only obtained the helix of the ear and spectacle frames
Chen and Bhanu [45]	Fused complexion from colour images and edges from a range of images	87.71	Fused complexion from colour images and edges from a range of images to perform ear detection
Prakash and Gupta [52]	Complexion and template-based technique	94	Used complexions and template-based techniques for automatic ear detection
Basrur et al. [53]	Gabor jets and reconstructed jets obtained via principal component analysis	NA	Introduced the notion of "jet space similarity," but did not report their algorithm's accuracy
Rahman et al. [54]	Cascaded AdaBoost technique supported Haar features	100	This system is widely known within the domain of face detection because of the Viola–Jones method, and it is a speedy and comparatively robust face detection technique
Chang et al. [55]	Multimodal recognition system	90.9	This system supported both the face and ear recognition
Naseem et al. [56]	General classification algorithm	98	This system investigated two crucial issues: feature extraction and robustness to occlusion
Nanni and Lumini [57]	Multi-matcher-based technique	NA	This system considers overlapping subwindows to extract local features
Yan and Bowyer [58]	Contour extraction algorithm	21	This system only used the ear contour using the active outline
Minaee et al. [59]	Independent component analysis and a radial basis function	94.11	The original ear image database and decomposing it into linear combinations of many basic images
Abdel-Mottaleb and Zhou [43]	Support vector machine	100	This approach is used for 3D ear detection and then a sliding window approach and linear SVM classifier to identify the ear

widely known within the domain of face detection because of the Viola–Jones method. It is a speedy and comparatively robust face detection technique. They trained the AdaBoost classifier to detect the ear region even in the presence of occlusions and degradation in image quality. They reported a 100% detection performance on the cascaded detector tested against 203 profile images from the UND database, with a false detection rate of 5×10. A second experiment detected 54 ears out of 104 partially occluded images from the XM2VTS database.

Chang et al. [55] built a multimodal recognition system that supported face and ear recognition. The manually identified coordinates of the triangular fossa and the antitragus are used for ear detection for the ear images. Their ear recognition system was supported by Eigen-ears' concept, using principal component analysis (PCA). They reported performance of 72.7% for the ear in one experiment, compared to 90.9% for the multimodal system, using 114 subjects from the UND, collection *E*, database.

Naseem et al. [56] proposed a general classification algorithm for (image-based) visual perception, supported by a sparse representation computed by L1 minimisation. This framework provides new insights into ear recognition's two crucial issues: feature extraction and robustness to occlusion. The ear portion is manually cropped from each image, and no normalisation of the ear region is required. They conducted several experiments using the UND and USTB databases with session variability, various head rotations, and different lighting conditions. These experiments yielded a high recognition rate within the order of 98%.

Nanni and Lumini [57] have proposed a multi-matcher-based technique for ear recognition that obtains the ear's appearance-based local properties. It considers overlapping subwindows to extract local features using Gabor filters. Further, Laplacian Eigen Maps are accustomed to reduce the feature vectors' dimensionality. The ear is represented using the features obtained from a group of the most discriminative subwindows selected using the sequential forward floating selection (SFFS) algorithm. Matching during this technique is performed by combining the outputs of several 1-nearest neighbour classifiers constructed on different subwindows. Another technique that supports the fusion of colour spaces is proposed by Nanni and Lumini, where few colour spaces are selected using the SFFS algorithm, and Gabor features are extracted from them. Matching is

Table 5: Summary of the ear algorithms using CNN.

Author	Dataset	Accuracy	Summary
Emeršič et al. [60]	NA	30	It used handcrafted feature extraction methods such as LBP, POEM, and CNN to obtain the ear identification
Tian et al. [21]	AMI, WPUT, IITD, and UERC	70.58, 67.01, 81.98, and 57.75	This system used deep CNN to perform ear recognition. There were occlusions like no earrings, headsets, or similar occlusions
Raveane et al. [64]	NA	98	This system used variable conditions due to the odd shape human ear and changing lighting conditions
Zhang and Mu [65]	UND and UBEAR	100 and 98.22	This system contained large occlusions, scale, and pose variation
Kohlakala and Coetzer [66]	AMI and IIT-Delhi	99.2 and 96.06	It is used to classify ears either in the foreground or background of the image. The binary contour image applied the matching for feature extraction, and this was performed by implementing Euclidean distance measure, which had a ranking to verify for authentication
Tomczyk and Szczepaniak [67]	NA	NA	It shows the published experimental results that the approach did the rotation equivalence property to detect rotated structures
Alshazly et al. [68]	Three ear datasets but not stated	22	The paper took seven performing handcrafted descriptors to extract the discriminating ear image. Then took the extracted ear and trained it using SVM to learn a suitable model
Alkababji and Mohammed [69]	NA	97.8	It used the PCA and a genetic algorithm for feature reduction and selection
Jamil et al. [70]	Very underexposed or overexposed database	97	This work was the first to test the performance of CNN on very underexposed or overexposed images
Hansley et al. [71]	UERC challenge	NA	This was performed using handcrafted descriptors, which were fused to improve recognition

administered by combining several nearest neighbour classifiers constructed on different colour components.

Yan and Bowyer [58] developed an automatic ear contour extraction algorithm. This was carried out by detecting the ear pit based on the position of the nose and cutting the ear contour using the active outline starting around the ear tip. This paper's results showed that 21% of the images tested were incorrectly segmented, but if they changed it to use only depth information and not colour, only 15% of the images were incorrectly segmented. A hybrid system for ear recognition was investigated by Minaee et al. [59]. This system combines an independent component analysis (ICA) and a radial basis function (RBF) network. This was conducted by taking the original ear image database and decomposing it into linear combinations of many basic images. Then, the corresponding coefficients of these combinations are used in the RBF network. They achieved 94.11% using two databases of segmented ear images.

A 3D ear detection system was investigated by Abdel-Mottaleb and Zhou [43]. They showed a novel shape-based feature set called histograms of categorised shapes (HCS). This approach is used for 3D ear detection and then a sliding window approach and linear support vector machine (SVM) classifier to identify the ear. They reported a perfect detection rate, a 100% detection rate, and a 0% false-positive rate.

5. Review of Ear Algorithms Using CNN

This section presents different algorithms using CNN used for ear recognition. This paper presents a description of these algorithms and suggests the most effective approach. A brief description of the ear algorithms using CNN is highlighted in Table 5.

Emeršič et al. [60] organized the dataset of the UERC. It was introduced and used for the benchmark, training, and testing sets. In this study, it was seen that handcrafted feature extraction methods such as linear binary pattern (LBP) [61], patterns of oriented edge magnitudes (POEM) [62], and CNN-based feature extraction methods were used to obtain the ear identification. In this challenge, one method needs to figure out a way to remove occlusions like earrings, hair, other obstacles, and background from the ear image. The occlusion was carried out by creating a binary ear mask, and then the system recognition was conducted using the handcrafted features. Another proposed approach was to calculate the score of matrices from the CNN-based features and handcrafted features when they are fused. A 30% detection rate was produced.

Tian et al. [21] applied a deep convolutional neural network (CNN) to ear recognition in which they designed a CNN—it was made up of three convolutional layers, a fully connected layer, and a softmax classifier. The database used was USTB ear, which consisted of 79 subjects with various pose angles. There were occlusions like no earrings, headsets, or similar occlusions. Chowdhury et al. [63] proposed an ear biometric recognition system that uses local features of the ear and then uses a neural network to identify the ear. The method estimates where the ear could be in the input image and then gets the edge features from the identified ear. After identifying the ear, a neural network matches the extracted feature with a feature database. The databases used in this system were AMI, WPUT, IITD, and UERC, which achieved an accuracy of 70.58%, 67.01%, 81.98%, and 57.75%, respectively.

Table 6: Sources of the articles reviewed.

S no.	Article source	Quantity
Conference		
1	IEEE	30
2	MPDI Applied Science	6
3	IET	4
4	CiteSeerX	2
Journal		
1	Scientific reports	3
2	SAIEE Africa Research Journal	1
3	Indonesian Journal of Electrical Engineering and Computer Science	2
4	ArXiv	9
5	ACM	3
6	ScienceDirect	9
7	Springer	17
8	IJESC	2
Books		
1	Manning	1
Total		89

Table 7: Differences between this review article and the recent/existing review papers.

Author(s) and date of publication	Paper title	Aim/focus/objective	Paper coverage (year) and scope
(1) This paper	Ear Biometrics using Deep Learning: A Survey	This paper proved that using a bag of feature techniques and the classification technique of deep learning using convolutional neural network was better than standard machine learning techniques	Eighty-nine (89) application papers that are deep learning ear identification methods are reviewed in this paper
(2) Emeršič et al. [60] 29 June 2017	Training convolutional neural networks with limited training data for ear recognition in the wild	It was a handcrafted feature extraction method, such as LBP and patterns of oriented edge magnitudes (POEM), and CNN-based feature extraction methods were used to obtain the ear identification	Forty-one (41) application papers that are deep learning ear identification methods are reviewed in this paper
(3) Tian et al. [21] 16 February 2017	Ear recognition based on deep convolutional network	This system used deep convolutional neural network (CNN) to ear recognition. There were occlusions like no earrings, headsets, or similar occlusions	Fifteen (15) application papers that are deep learning ear identification methods are reviewed in this paper
(4) Raveane et al. [64] 18 June 2019	Ear detection and localization with convolutional neural networks in natural images and videos	This system used variable conditions, and this could also be because of the odd shape of the human ears and changing lighting conditions	Thirty-five (35) application papers that are deep learning ear identification methods are reviewed in this paper

Raveane et al. [64] presented that it is difficult to precisely detect and locate an ear within an image. This challenge increases when working with variable conditions, and this could also be because of the odd shape of the human ears and changing lighting conditions. The changing profile shape of an ear when photographed is displayed [64]. The ear detection system was a multiple convolutional neural network with a detection grouping algorithm to identify the ear's presence and location. The proposed method matches other methods' performance when analysed against clean and purpose-shot photographs, reaching an accuracy of upwards of 98%. It outperforms other works with a rate of over 86% when the system is subjected to noncooperative natural images where the subject appears in challenging orientations and photographic conditions.

Multiple scale faster region-based convolutional neural network (Faster R-CNN) to detect ears from 2D profile images was proposed by Zhang and Mu [65]. This method uses three regions of different scales to detect information from the ears' location within the context of the ear image. The system was tested with 200 web images and achieved an accuracy of 98% . Other experiments conducted were on the Collection J2 of the University of Notre Dame Biometrics Database (UND-J2) and the University of Beira Interior Ear (UBEAR) dataset; these achieved a detection rate of 100% and 98.22%, respectively, but these datasets contained large occlusions, scale, and pose variation.

Kohlakala and Coetzer [66] presented semiautomated and fully automated ear-based biometric verification systems. A convolutional neural network (CNN) and

Table 8: Cont. differences between this review article and the recent/existing review papers.

Author(s) and date of publication	Paper title	Aim/focus/objective	Paper coverage (year) and scope
(5) Zhang and Mu [65] 24 January 2017	Ear detection under uncontrolled conditions with multiple scale faster region-based convolutional neural networks	This system contained large occlusions, scale, and pose variation	Forty-one (41) application papers that are deep learning ear identification methods are reviewed in this paper
(6) Kohlakala and Coetzer [66] 1 June 2021	Ear-based biometric authentication through the detection of prominent contour	It is used to classify ears either in the foreground or background of the image. The binary contour image applied the matching for feature extraction, and this was carried out by implementing Euclidean distance measure, which had a ranking to verify for authentication	Twenty-one (21) application papers that are deep learning ear identification methods are reviewed in this paper
(7) Tomczyk and Szczepaniak [67] 13 December 2019	Ear detection using convolutional neural network on graphs with filter rotation	It shows the published experimental results that the approach performed the rotation equivalence property to detect rotated structures	Forty (40) application papers that are deep learning ear identification methods are reviewed in this paper
(8) Alshazly et al. [68] 8 December 2019	Handcrafted versus CNN features for ear recognition	The paper took seven performing handcrafted descriptors to extract the discriminating ear image. They then took the extracted ear and trained it using Support Vector Machines (SVM) to learn a suitable model	Seventy-three (73) application papers that are deep learning ear identification methods are reviewed in this paper

Table 9: Cont. differences between this review article and the recent/existing review papers.

Author(s) and date of publication	Paper title	Aim/focus/objective	Paper coverage (year) and scope
(9) Alkababji and Mohammed [69] 1 April 2021	Real-time ear recognition using deep learning	It used the principal component analysis (PCA) and a genetic algorithm for feature reduction and selection	Twenty-three (23) application papers that are deep learning ear identification methods are reviewed in this paper
(10) Jamil et al. [70] 1 August 2018	Can convolution neural network (CNN) triumph in ear recognition of uniform illumination invariant?	They considered that their work was the first to test the performance of CNN on very underexposed or overexposed images	Thirty-two (32) application papers that are deep learning ear identification methods are reviewed in this paper
(11) Hansley et al. [71] 24 October 2017	Employing fusion of learned and handcrafted features for unconstrained ear recognition	This was conducted using handcrafted descriptors, which were fused to improve recognition	Thirty-one (31) application papers that are deep learning ear identification methods are reviewed in this paper

morphological postprocessing were used to manually identify the ear region. They are used to classify ears either in the foreground or background of the image. The binary contour image applied the matching for feature extraction, and this was carried out by implementing a Euclidean distance measure, which had a ranking to verify for authentication. The Mathematical Analysis of Images ear database and the Indian Institute of Technology, Delhi, ear database were two databases, which achieved 99.20% and 96.06%, respectively.

Geometric deep learning (GDL) generalises convolutional neural network (CNN) to non-Euclidean domains, presented by [67] Tomczyk and Szczepaniak. It used convolutional filters with a mixture of Gaussian models. These filters were used so that the images could be easily rotated without interpolation. Their paper published experimental results on the approach of the rotation equivalence property to detect rotated structures. The result showed that it did not require labour-intensive training on all rotated and non-rotated images.

Alshazly et al. [68] presented and compared ear recognition models built with handcrafted and convolutional neural networks (CNN) features. The paper took seven handcrafted descriptors to extract the discriminating ear image. The extracted ear was trained using Support Vector Machines (SVM) to learn a suitable model, after which the CNN-based model used the AlexNet architecture. The results obtained on three ear datasets show the CNN-based models' performance by 22%. This paper also investigated if the left and right ears have symmetry. The results obtained by the two datasets indicate a high impact of balance between the ears.

Alkababji and Mohammed [69] presented the use of a deep learning item detector, which they called faster region-based convolutional neural networks (Faster R-CNN) for ear detection. This convolutional neural network (CNN) is used for feature extraction. It used Principal Component Analysis (PCA) and a genetic algorithm for feature reduction and selection. It also used a connected artificial neural network as the matcher. The results achieved an accuracy of 97.8% success.

Jamil et al. [70] built and trained a CNN model for ear biometrics in various uniform illuminations measured using lumens. They considered that their work was the first to test the performance of CNN on very underexposed or over-exposed images. The results showed that for images with uniform illumination and a luminance of above 25 lux, the results achieved were 100%. The CNN model had problems recognising images when the lux was below ten, but still obtained an accuracy of 97%. This result shows that the CNN architecture performs just as well as the other systems. It was found that the data set had rotations that affected the results.

Hansley et al. [71] presented an unconstrained ear recognition framework that was better than the current state-of-the-art systems using publicly available databases. They developed CNN-based solutions for ear normalisation and description. This was performed using handcrafted descriptors, which were fused to improve recognition, and was carried out in two stages. The first stage was to find the landmark detectors, which were untrained scenarios. The next step was to generate a geometric image normalisation to boost the performance. It was seen that the CNN descriptor was better than other CNN-based works in the literature. The obtained results were higher than different reported results for the UERC challenge.

6. Difference between Reviewed Articles

Tables 6 and 7 show the comparison of this review paper with recent/existing review papers to establish their differences. A critical analysis of Tables 6 and 7 reveals that the most recent and closest review paper to this article is the excellent review work. Tables 8 and 9 show the differences between the review article and the existing review papers.

7. Conclusion

This paper presented a comparative survey of various convolutional neural network architectures, with their strengths and weaknesses. A thorough analysis of the existing deep convolutional neural network methods used for ear identification was discussed. Furthermore, the paper discussed and investigated the success of using the ear as a primary biometric system for identification and verification. It was found that other works battled to identify the ear if pose and angle of the image were changed. This will be looked at in the future as to how this can be eliminated. Also, it was found that if clothes, hair, ear ornaments, and jewellery were not removed, it interfered with the identification of an ear. In addition, a study was performed on ear identification benchmarks and their performance on other CNN models measured by standard evaluating metrics.

Future work will be to investigate and implement EfficientNet models to automatically identify ears on the most prominent and publicly available datasets. EfficientNets that achieved state-of-the-art performance over other architectures to maximize accuracy and efficiency were explored and fine-tuned on profile images. The fine-tuning technique is valuable to utilize rich generic features learned from significant dataset sources such as ImageNet to complement the lack of annotated datasets affecting the ear domains.

Abbreviations

NN:	Neural network
CNN:	Convolutional neural network.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Comput Surv*, vol. 45, no. 2, pp. 1–35, 2013.

[2] B. C. Bir, *Ear Biometrics*, Springer, Boston, MA, USA, 3D edition, 2009.

[3] J. Heaton, I. Goodfellow, B. Yoshua, and C. Aaron, *Deep Learning*, Springer, New York, NY, USA, 2018.

[4] F. Chollet, *Deep learning with Python*, Vol. 361, Manning, New York, NY, USA, 2018.

[5] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, June 2015.

[6] J. Dai, K. He, and J. Sun, "Boxsup: exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1635–1643, 2015.

[7] A. Lomuscio and L. Maganti, "An Approach to reachability Analysis for feed-forward relu neural networks," 2017, https://arxiv.org/abs/1706.07351?context=cs.

[8] J. Kleesiek, G. Urban, A. Hubert et al., "Deep MRI brain extraction: a 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, 2016.

[9] F. Bonanno, G. Capizzi, G. L. Sciuto, C. Napoli, G. Pappalardo, and E. Tramontana, "A cascade neural network architecture investigating surface plasmon polaritons propagation for thin metals in openmp," in *International Conference on Artificial Intelligence and Soft Computing-*Springer, New York, NY, USA, 2014.

[10] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "NAS-Unet: neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, Article ID 44247, 2019.

[11] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, and M. S. Nasrin, "The history Began from Alexnet: A comprehensive survey on deep Learning Approaches," 2018, https://arxiv.org/abs/1803.01164.

[12] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Ensembles of deep learning models and transfer learning for ear recognition," *Sensors*, vol. 19, no. 19, p. 4139, 2019.

[13] R. R. Hallac, J. Lee, M. Pressler, J. R. Seaward, and A. A. Kane, "Identifying ear abnormality from 2D photographs using convolutional neural networks," *Scientific Reports*, vol. 9, no. 1, Article ID 18198, 2019.

[14] E. Ž, D. Štepec, V. Štruc, P. Peer, A. George, and A. Ahmad, "The unconstrained ear recognition challenge," in *Proceedings of the 2017 IEEE International joint Conference on Biometrics (IJCB)*, pp. 715–724, Denver, CO, USA, October 2017.

[15] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Deep convolutional neural networks for unconstrained ear recognition," *IEEE Access*, vol. 8, Article ID 170295, 2020.

[16] Y. Zhang, Z. Mu, L. Yuan, and C. Yu, "Ear verification under uncontrolled conditions with convolutional neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 185–198, 2018.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, Las Vegas, NV, USA, June 2016.

[18] Z. Li, Z. Hu, J. Xu, T. Tan, H. Chen, and Z. Duan, "Computer-aided diagnosis of lung carcinoma using deep learning-a pilot study," 2018, https://arxiv.org/abs/1803.05471.

[19] K. Radhika, K. Devika, T. Aswathi, P. Sreevidya, V. Sowmya, and K. Soman, "Performance analysis of NASNet on unconstrained ear recognition," in *Nature Inspired Computing for Data Science*Springer, New York, NY, USA, 2020.

[20] Y. Li, "Deep reinforcement Learning: An overview," 2017, https://arxiv.org/abs/170107274.

[21] L. Tian and Z. Mu, "Ear recognition based on deep convolutional network," in *Proceedings of the 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 437–441, Datong, China, October 2016.

[22] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, "Dermoscopic image segmentation via multistage fully convolutional networks," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2065–2074, 2017.

[23] S. Dodge, J. Mounsef, and L. Karam, "Unconstrained ear recognition using deep neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 207–214, 2018.

[24] R. Hussain, A. Lalande, K. B. Girum, C. Guigou, and A. Bozorg Grayeli, "Automatic segmentation of inner ear on CT-scan using auto-context convolutional neural network," *Scientific Reports*, vol. 11, no. 1, pp. 4406–4410, 2021.

[25] S. Zhou, F. Wang, Z. Huang, and J. Wang, "Discriminative feature learning with consistent attention regularization for person re-identification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8040–8049, Seoul, Republic of Korea, October 2019.

[26] A. Kumar, "Iit delhi ear database version 1.0," 2007, http://webold.iitd.ac.in/biometrics/Database_Ear.htm.

[27] R. Raposo, E. Hoyle, A. Peixinho, and H. ProenÃ§a, "UBEAR: A dataset of ear images captured on-the-move in uncontrolled conditions," in *Proceedings of the 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, Paris, France, April 2011.

[28] Ž Emeršič, V. Štruc, and P. Peer, "Ear recognition: more than a survey," *Neurocomputing*, vol. 255, pp. 26–39, 2017.

[29] V. T. Hoang, "EarVN1.0: a new large-scale ear images dataset in the wild," *Data in Brief*, vol. 27, Article ID 104630, 2019.

[30] Y. Zhang, Z. C. Mu, L. Yuan, C. Yu, and L Qing, "USTB-Helloear: a large database of ear images photographed under uncontrolled conditions," in *Image and Graphics*, Springer, New York, NY, USA, 2017.

[31] E. Gonzalez, L. Alvarez, and L. Mazorra, "Ami Ear Database," 2012, http://ctim.ulpgc.es/research_works/ami_ear_database/ .

[32] D. Frejlichowski and N. Tyszkiewicz, "The west pomeranian university of technology ear database - a tool for testing biometric algorithms," *Image Analysis and Recognition*, Springer, Berlin, Germany, 2010.

[33] V. Emeršič and P. Peer, "Ear Biometric database in the wild," in *Proceedings of the 2015 4th International Work Conference on Bioinspired Intelligence (IWOBI)*, pp. 27–32, San Sebastian, Spain, June 2015.

[34] M. A. Carreira-Perpinan, "Compression neural networks for feature extraction: Application to human recognition from ear images," Master's thesis, Faculty of Informatics, Technical University of Madrid, Madrid, Spain, 1995.

[35] S. Prakash, U. Jayaraman, and P. Gupta, "Connected component based technique for automatic ear detection," in *Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 2741–2744, IEEE, Cairo, Egypt, November 2009.

[36] I. Alberink and A. Ruifrok, "Performance of the FearID earprint identification system," *Forensic Science International*, vol. 166, no. 2-3, pp. 145–154, 2007.

[37] P. Yan and K. Bowyer, "Empirical evaluation of advanced ear biometrics," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, p. 41, San Diego, CA, USA, September 2005.

[38] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.

[39] T. Sim, S. Baker, and M. Bsat, *The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces*, Carnegie Mellon University, Pittsburgh, PA, 2001.

[40] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: the extended M2VTS database," in *Proceedings of the Second International Conference on Audio and Video-Based Biometric Person Authentication*, pp. 965-966, 1999.

[41] A. Abaza, *High performance image processing techniques in Automated identification systems*, West Virginia University, Morgantown, WV, USA, 2008.

[42] S. Ansari and P. Gupta, "Localization of ear using outer helix curve of the ear," in *Proceedings of the 2007 International Conference on Computing: Theory and Applications (ICCTA'07)*, pp. 688–692, IEEE, Kolkata, India, March 2007.

[43] M. Abdel-Mottaleb and J. Zhou, "Human ear recognition from face profile images," in *International Conference on Biometrics*Springer, New York, NY, USA, 2006.

[44] T. Yuizono, Y. Wang, K. Satoh, and S. Nakayama, "Study on individual recognition for ear images by using genetic local search," in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600)*, pp. 237–242, IEEE, Honolulu, HI, USA, May 2002.

[45] H. Chen and B. Bhanu, "Human ear recognition in 3D," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 718–737, 2007.

[46] B. Arbab-Zavar and M. S. Nixon, "On shape-mediated enrolment in ear biometrics," in *International Symposium on Visual Computing*, Springer, New York, NY, USA, 2007.

[47] M. Burge and W. Burger, *Ear Biometrics*, pp. 273–285, Biometrics Springer, New York, NY, USA, 1996.

[48] M. Choras, "Image feature extraction methods for ear biometrics–a survey," in *Proceedings of the 6th International Conference on Computer Information Systems and Industrial Management Applications (CISIM'07)*, June 2007.

[49] D. Shailaja and P. Gupta, "A simple geometric approach for ear recognition," in *Proceedings of the 9th International Conference on Information Technology (ICIT'06)*, pp. 164–167, IEEE, Bhubaneswar, India, December 2006.

[50] A. H. Cummings, M. S. Nixon, and J. N. Carter, "A novel ray analogy for enrolment of ear biometrics," in *Proceeding of the 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–6, Washington, DC, USA, September 2010.

[51] A. Abaza, C. Hebert, and M. A. F. Harrison, "Fast learning ear detection for real-time surveillance," in *Proceedings of the 2010 fourth IEEE international Conference on Biometrics: theory, Applications and systems (BTAS)*, pp. 1–6, Washington, DC, USA, September 2010.

[52] S. Prakash and P. Gupta, "An efficient ear localization technique," *Image and Vision Computing*, vol. 30, no. 1, pp. 38–50, 2012.

[53] V. Basrur, F. Yang, T. Kushimoto et al., "Proteomic analysis of early melanosomes: identification of novel melanosomal proteins," *Journal of Proteome Research*, vol. 2, no. 1, pp. 69–79, 2003.

[54] M. Rahman, M. R. Islam, N. I. Bhuiyan, B. Ahmed, and M. A. Islam, "Person identification using ear biometrics," *International Journal of The Computer, the Internet and Management*, vol. 15, no. 2, pp. 1–8, 2007.

[55] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1160–1165, 2003.

[56] I. Naseem, R. Togneri, and M. Bennamoun, "Sparse representation for ear biometrics," in *International Symposium on Visual Computing*, Springer, New York, NY, USA, 2008.

[57] L. Nanni and A. Lumini, "A multi-matcher for ear authentication," *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2219–2226, 2007.

[58] P. Yan and K. W. Bowyer, "Biometric recognition using 3D ear shape," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1297–1308, 2007.

[59] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang, "Biometric recognition using deep learning: a survey," 2019, https://arxiv.org/abs/1912.00271.

[60] Z. Emeršič, D. Štepec, V. Štruc, and P. Peer, "Training convolutional neural networks with limited training data for ear recognition in the wild," in *Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Washington, DC, USA, May 2017.

[61] Zq Wang and Xd Yan, "Multi-scale feature extraction algorithm of ear image," in *Proceedings of the 2011 International Conference on Electric Information and Control Engineering*, pp. 528–531, IEEE, Wuhan, China, April 2011.

[62] N. S. Vu, H. M. Dee, and A. Caplier, "Face recognition using the POEM descriptor," *Pattern Recognition*, vol. 45, no. 7, pp. 2478–2488, 2012.

[63] D. P. Chowdhury, S. Bakshi, G. Guo, and P. K. Sa, "On applicability of tunable filter bank based feature for ear biometrics: a study from constrained to unconstrained," *Journal of Medical Systems*, vol. 42, no. 1, pp. 11–20, 2018.

[64] W. Raveane, P. L. Galdamez, and M. A. Gonzalez Arrieta, "Ear detection and localization with convolutional neural networks in natural images and videos," *Processes*, vol. 7, no. 7, p. 457, 2019.

[65] Y. Zhang and Z. Mu, "Ear detection under uncontrolled conditions with multiple scale faster region-based convolutional neural networks," *Symmetry*, vol. 9, no. 4, p. 53, 2017.

[66] A. Kohlakala and J. Coetzer, "Ear-based biometric authentication through the detection of prominent contours," *SAIEE Africa Research Journal*, vol. 112, no. 2, pp. 89–98, 2021.

[67] A. Tomczyk and P. S. Szczepaniak, "Ear detection using convolutional neural network on graphs with filter rotation," *Sensors*, vol. 19, no. 24, p. 5510, 2019.

[68] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Handcrafted versus CNN features for ear recognition," *Symmetry*, vol. 11, no. 12, p. 1493, 2019.

[69] A. M. Alkababji and O. H. Mohammed, "Real time ear recognition using deep learning," *Telkomnika*, vol. 19, no. 2, pp. 523–530, 2021.

[70] N. Jamil, A. Almisreb, S. M. Z. S. Z. Ariffin, N. Md Din, and R. Hamzah, "Can Convolution neural network (CNN) triumph in ear recognition of uniform illumination invariant?" *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 11, 2018.

[71] E. E. Hansley, M. P. Segundo, and S. Sarkar, "Employing fusion of learned and handcrafted features for unconstrained ear recognition," *IET Biometrics*, vol. 7, no. 3, pp. 215–223, 2018.