A Noble Approach to 2D Ear Recognition System using Hybrid Transfer Learning

Ravishankar Mehta
Machine Vision and Intelligence Lab
National Institute of Technology
Jamshedpur, India
Rmehta.online@gmail.com

*Akbar Sheikh-Akbari
Leeds Beckett University
School of Built, Env. Eng. & Computing
Leeds, U.K.
*A.Sheikh-Akbari@leedsbeckett.ac.uk

Koushlendra K. Singh
Machine Vision and Intelligence Lab
National Institute of Technology
Jamshedpur, India
koushlendra.cse@nitjsr.ac.in

***Abstract*—Convolutional Neural Networks (CNNs) have emerged as a popular choice of researchers for their robust feature extraction and information mining capability. In the last decades, CNNs have depicted impressive performance on various applications of computer vision tasks like object detection, image segmentation, and image classification. As a consequence, the ear-based recognition system has not gained many benefits from deep learning and CNN-based applications and is still lacking behind due to the availability of sufficient data and varying conditions of captured sample images. In this paper, transfer learning techniques have been applied to the well-known convolutional neural network model VGG16 integrated with the support vector machine(SVM) that acts as a hybrid algorithm for recognizing the person using their ear images. The proposed model is validated on an ear dataset containing a total of 2600 images with variability in terms of pose, rotation, and illumination changes. The proposed model is able to classify the ear images with the highest recognition accuracy of 98.72%. To show the effectiveness of the proposed model, comparative studies of the proposed model with other existing methods have been reported in the literature.**

***Index Terms*—Transfer learning, Deep learning, Ear recognition, Feature extraction**

I. Introduction

Research in the field of ear-based recognition systems getting more and more popular over recent years mainly due to their potential applications in forensics, security, monitoring, and surveillance. Various studies have been conducted by researchers to build a recognition system that uses human ear images [1]. Machine learning-based ear recognition systems perform the classification tasks in a single domain by training each classifier individually [2]. In contrast, transfer learning approaches utilize cross-domain learning [3]. The major concern behind the use of transfer learning is to transfer the skill learned in one domain to the other related domains. While developing the transfer learning-based model, one of the major challenges that are to be dealt with is to ensure the transfer of positive knowledge since the transfer of negative or wrong knowledge can lead to the degradation of the system's performance. In the proposed method, a transfer learning technique has been applied for recognizing the person using ear images. The major contributions of this paper are as follows:

1) We have conducted a strategy to fine-tune the VGG16 network with fewer trainable parameters to improve system accuracy.
2) For classification tasks, we have utilized the support vector machine as a classifier to classify the features extracted by transfer learning-based model using VGG16.
3) An extensive experiment has been conducted to evaluate the performance of the proposed model and is compared with the state-of-the-art methods.

The structure of the paper is organized as follows: Section II discusses the technique and architecture of various deep learning models. The proposed methodology and dataset descriptions are introduced in section III. The experimental setup and results are presented in section IV. Finally, the paper is concluded in section V.

II. Related Work

Over the past decade, ear, biometric modalities, has gained the attention of researchers. Due to high inter and intra-class variation in ear image data, hand-crafted features fails in large-scale datasets. In this regard, deep CNNs have great success in the field of computer vision and machine learning. The robust feature extraction mechanisms of deep convolutional neural networks facilitate the researcher to apply them in a variety of applications. In this context, transfer learning has shown tremendous performance that utilizes pre-trained networks (e.g. VGG16, VGG19) to extract the features by fine-tuning the network weights through training the model with the new dataset. Initially, the AlexNet model has been proposed to handle object recognition tasks [4]. However, the major problem with such CNN based model is the training of this model. To train the network, the ILSVRC dataset that uses an augmentation technique has been proposed [5]. For further improvement of recognition accuracy, researchers applied deeper CNN architectures in the field of computer vision tasks. Since the accuracy of CNN based model highly relies on the availability of a large dataset, highly computational unit, and depth of the network. The first requirement of such a model is solved with the availability of the ILSVRC dataset which is publically available. The availability of highly configured GPU units could solve the second requirement. But the requirement of the last constraint i.e. measuring the depth of the network is uncertain since there is no such measure that can limit the network depth. The deeper networks can extract more robust and complex features. Simonyan et. al. proposed

VGG16 architecture to meet the requirement and solves many image classification tasks in the field of computer vision [6]. In contrast to AlexNet architecture, the VGG16 network uses the replicative structure of convolution, relu, and pooling layer and an increased number of such layers to build a deeper network. Further improvement in VGG16 architecture has been reported in VGG19 which overcomes the drawback of AlexNet and improves the system accuracy. While deep learning-based models show extremely good performance with biometrics such as face, iris, and fingerprint, their applications in ear recognition are still lacking behind due to the availability of large-scale datasets [7]. Some of the popular approaches that employ deep learning for ear recognition are discussed in [7] [8] [9] [10]. Galdámez et al. were the first who applied deep learning using CNN for ear recognition of video images taken in a controlled environment [10]. Since video streaming requires recognition to be conducted rapidly, a highly optimized CNN architecture is needed. Though their approach shows encouraging results, their network training is done in a non-collaborative environment of controlled images.

Much research has been conducted using transfer learning for various applications like disease classification [11], text classifications [12], sentimental classifications [13], link prediction [14], rank learning [15]. In this paper, the transfer learning technique using the VGG16 model for feature extraction and support vector machine for classification purposes has been utilized for ear recognition.

III. Materials and Methodology

In this section, the description of the ear image dataset and a detailed explanation of VGG16 network architecture is presented. The fine-tuning strategy has been employed in the pre-trained VGG16 network and the support vector machine is used to perform the classification based on features learned by VGG16 model.

A. Proposed model using VGG16 and SVM

A general deep learning-based architecture mainly consists of pre-processing, feature extractions, and classification components. The proposed architecture consists of a VGG16 model and SVM as a classifier, as shown in Figure 1.

In this paper, we have fine-tuned the network of the VGG16 model over the ear dataset [16]. It is divided into two major steps. In the first step, we extracted the intrinsic features of ear images through a transfer learning approach with the help of VGG16, and in the second step, classification is performed with the help of SVM. The entire method from feature extraction to classifications is comprised of pre-processing, feature extraction, fine-tuning, and classifications.

As depicted in Figure 1, the input image is passed through a series of convolutional layers, each followed by a max pooling layer. The convolutional layers are responsible for detecting and extracting features from the input image. The max pooling layers reduce the spatial dimensionality of the feature maps, making the network more efficient. After the last max pooling layer, the feature maps are flattened into a vector and passed through an SVM layer. The SVM layer is trained to classify the input image into one of several classes based on the features extracted by the earlier convolutional layers. Overall, this architecture is commonly used in image classification tasks, where the SVM layer is used to make the final classification decision based on the features learned by the CNN layers.

B. Dataset Description

In this paper, a total of 2600 ear images of 13 distinct persons have been utilized to validate the performance of the proposed model [16]. This dataset contains color as well as black-and-white images of the human ear. The dataset is downloaded from Kaggle whose link is provided in the reference section. We split the dataset into train and test sets in the ratio of 9:1. Among the 2600 images, a total of 2340 images are used for training the network, and 260 images are used to test the network.

C. VGG16 Architecture and its Fine Tuning

In this paper, VGG 16 network architecture has been used for feature extraction tasks. This network improves the recognition accuracy on challenging image datasets which are unconstrained [17] [18] [19]. It replaces the large kernel-sized filters of the first and second convolution layers with a multiple of 3x3 filters. For evaluating the proposed method, the input images are re-scaled and cropped into a size of 150x150. The earlier layers of the pre-trained VGG16 model are retained and fixed for extracting the features from ear images. While applying the concept of transfer learning in the presented work, we replaced the last three layers of fully connected layers of the VGG16 model with the set of layers that can classify 13 classes to recognize the 13 subjects. We have supplied a fully connected layer of filter size 128x128 that can adopt new output for 13 subjects. One Rectified Linear Unit (ReLU) activation function has been added to the network to solve the non-linearity problem. While this function makes the training procedure faster, it also prevents the model from vanishing gradient problems. One more fully connected layer is added with 13 output neurons to perform the classifications among 13 subjects. For speeding up the learning in the new layers than the transferred layer, the weights in the last fully connected layers are initialized to 13.

IV. Result and Discussions

For performance evaluation of the proposed model, the entire work is evaluated on 2600 ear images of 13 different classes. The dataset was split into the training and test sets with a ratio of 9:1. The training procedure is conducted on a system with a simple CPU having a 3.70 GHz processor with 16 GB Memory (RAM) using Python version 3.7.8. We have used TensorFlow for building the proposed model. The input image size is kept at 150x150. The top layer of the VGG16 model has been fine-tuned so that specific patterns from the input ear image can be learned. While performing the experiments, we have set different values to hyper-parameters as illustrated in Table I. We run the model for 30 epochs. The batch size has been set fixed to 16 and squared_hinge loss functions have

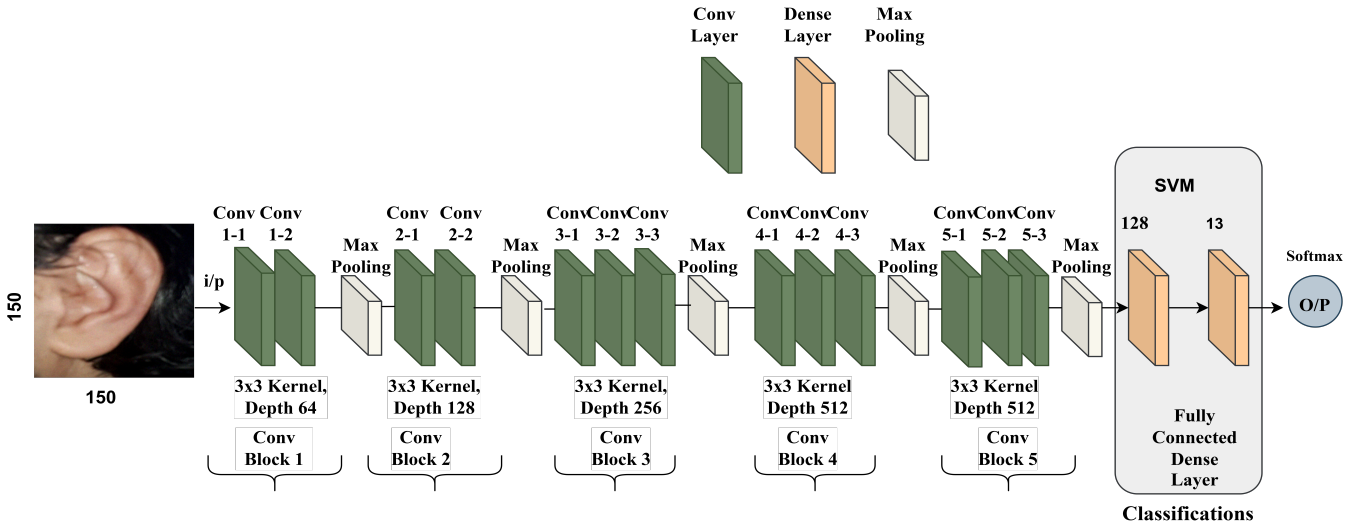

Fig. 1. Proposed model for ear recognition using transfer learning and SVM.

TABLE I
DIFFERENT VALUE OF HYPERPARAMETERS USED IN THE PROPOSED MODEL

Hyper Parameters	Values
Optimizer	Adam
Regularization	L2 = 0.01
Dropout	0.5
Batch Size	16
Learning Rate	0.001
Learning iterations	10
Filter Size	3x3
Loss function	squared_hinge
No. of epochs	30

been used for the entire experiment. A filter of size 3x3 has been used for the convolution operations. Adam optimizer has been used to tune the network parameters in an automated manner without the intervention of an operator. This optimizer prevents the model from over-fitting and optimizes the loss. For noisy environments, the Adam optimizer provides the ability to tackle the problem of sparse gradients. We have used Ridge regularization (L2) with a value of 0.01 in the last dense layer to control the strength of the penalty applied to the squared value of the model's weights. Determining the optimal value for the regularizer is a critical task. Since its higher value penalizes the model heavily due to the large weights thereby making the model less likely to overfit the training data. On the other hand, its lower value will be less penalized to the model thereby making the model more complex that is more likely to overfit the training data. It has been observed that regularizer L2 = 0.01 balances the trade-off between the model complexity and generalization performances. For the loss functions, we have used the squared_hinge function. This function makes the last layer of the model as SVM classifier. At the last layer, we used the softmax activation function since our problem is a multi-classification problem. For evaluating the performance of the proposed model, the accuracy in terms of validation loss and validation accuracy has been calculated and is depicted in Figure 2. Using SVM as a classifier in the last layer of transfer learning, the performance of the proposed model converges after 30 epochs and it achieves its best performance with a validation accuracy of 99.23% and a loss of 0.04. To measure the effectiveness of the SVM as the classifier, we conducted the experiments by excluding the SVM and using only a fully connected layer at the end. It has been observed that with a batch size of 16 and a learning rate of 0.001, the model starts converging after 25 epochs. At this time the model achieves performance accuracy of 98.71% and a loss of 0.09. Its performances have been plotted in Figure 3 which gives validation accuracy vs. epoch and validation loss vs. epochs.

To show the effectiveness of the proposed model the performance of the proposed model in terms of accuracy was compared with the other state-of-the-art methods and is illustrated in Table 2. From Table 2, it can be seen that the proposed model gives very good results in comparison with the existing models.

V. CONCLUSION

In this paper, the concept of the hybrid transfer learning approach with the help of VGG16 and SVM as a classifier for ear recognition was introduced For evaluating the performance of the model, we have conducted an experiment on 2600 ear images downloaded from Kaggle. First, we fine-tuned the pre-trained VGG16 model to recognize 13 classes only using 2340 images for training and 260 images for testing. The VGG16 pre-trained model was then used to extract the features from ear images. The fully connected layer of the VGG16 model was replaced with an SVM classifier for classification purposes. The dataset contains images of different variability in which our model achieves more than 99% accuracy. The performance of the proposed model was compared with the state-of-the-art transfer learning approaches such as AlexNet, GoogleNet, and MobileNet, in which the proposed model shows superior

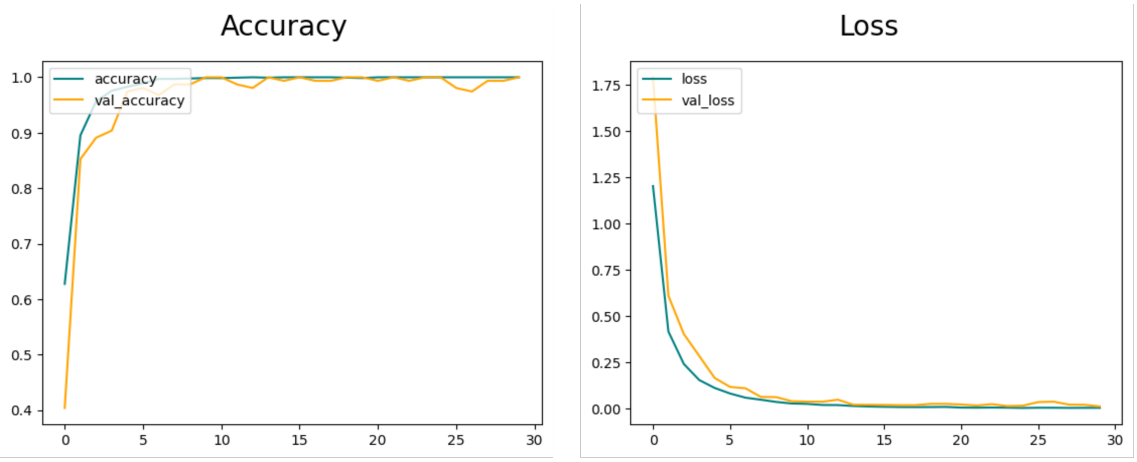

Fig. 2. Validation accuracy and validation loss against each epoch of the proposed model with SVM.

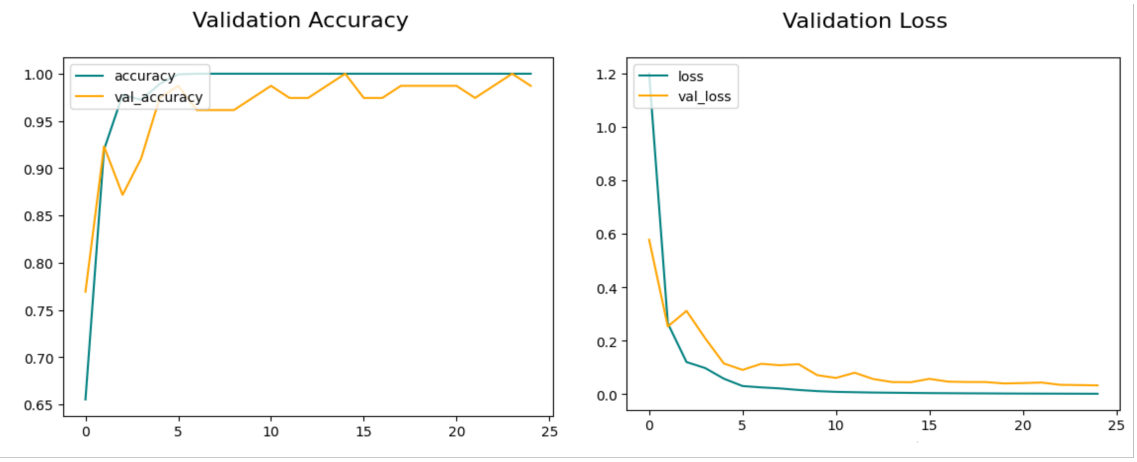

Fig. 3. Validation accuracy and validation loss against each epoch of the proposed model without SVM

TABLE II
PERFORMANCE COMPARISON OF THE PROPOSED MODEL WITH STATE-OF-THE-ART METHODS

Related Article	**Dataset**	**Accuracy** (%)
Transfer learning for image classification(AlexNet) [20], 2018	caltech256	87.08
Transfer learning for image classification(VGG16) [20], 2018	caltech256	88.04
Transfer learning for image classification(VGG19) [20], 2018	caltech256	88.63
Transfer learning for image classification(AlexNet) [20], 2018	GHIM10k	96.88
Transfer learning for image classification(VGG16) [20], 2018	GHIM10k	98.574
Transfer learning for image classification(VGG19) [20], 2018	GHIM10k	99.38
Domain adaptation for ear recognition using deep CNN(AlexNet) [21], 2018	Multi-PIE	96.71
Domain adaptation for ear recognition using deep CNN(VGG16) [21], 2018	Multi-PIE	98.57
Domain adaptation for ear recognition using deep CNN(GoogleNet) [21], 2018	Multi-PIE	97.80
Evaluation of Deep Learning Models for Ear Recognition(AlexNet) [22], 2019	UERC	94.67
Evaluation of Deep Learning Models for Ear Recognition(GoogleNet) [22], 2019	UERC	93.33
Evaluation of Deep Learning Models for Ear Recognition(MobileNet) [22], 2019	UERC	95.67
Wavelet-Based Multi-Band [23] , 2021	IITD-II	94.47
Ear recognition using six deep CNN model [24], 2021	IITD-II	97.36
Transfer Learning: A way for Ear Biometric Recognition(CNN) [25], 2022	IITD-II	88.37
Transfer Learning: A way for Ear Biometric Recognition(VGG16) [25], 2022	IITD-II	88.73
Transfer Learning: A way for Ear Biometric Recognition(ResNet50) [25], 2022	IITD-II	89.71
Proposed transfer learning model with SVM as classifier	**Kaggle**	**99.23**

performance over all these models. The concept of hybrid transfer learning can be applied to other real-time applications like human action recognition, and object detection with the efficacy of the features extractions techniques on video datasets.

REFERENCES

[1] Emeršič, Ž., Štruc, V. and Peer, P., 2017. Ear recognition: More than a survey. Neuro Computing, 255, pp.26-39.

[2] Yuan, L. and Zhang, F., 2009, July. Ear detection based on improved adaboost algorithm. In 2009 International Conference on Machine Learning and Cybernetics (Vol. 4, pp. 2414-2417). IEEE.

[3] Chowdhury, D.P., Bakshi, S., Pero, C., Olague, G. and Sa, P.K., 2022. Privacy Preserving Ear Recognition System Using Transfer Learning in Industry 4.0. IEEE Transactions on Industrial Informatics.

[4] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", Advances in neural information processing systems, pp. 1097-1105, 2012.

[5] J. D. A. Berg and L. Fei-Fei, Large scale visual recognition challenge 2010, 2010, [online] Available: http://image-net.org/download.

[6] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014.

[7] Alay, N. and Al-Baity, H.H., 2020. Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. Sensors, 20(19), p.5523.

[8] Ž. Emersic, D. Štepec, V. Štruc and P. Peer, Training convolutional neural networks with limited training data for ear recognition in the wild, 2017.

[9] Y. Zhang and Z. Mu, "Ear detection under uncontrolled conditions with Multiple Scale Faster Region-Based Convolutional Neural Networks", Symmetry, vol. 9, no. 4, pp. 1-19, 2017.

[10] Mehta, R. and Singh, K.K., 2023, January. Ear Recognition System Using Averaging Ensemble Technique.In Machine Learning, Image Processing, Network Security and Data Sciences: 4th International Conference, MIND 2022, Virtual Event, January 19–20, 2023, Proceedings, Part II, pp. 220-229.

[11] Saikia, T., Kumar, R., Kumar, D. and Singh, K.K., 2022. An Automatic Lung Nodule Classification System Based on Hybrid Transfer Learning Approach. SN Computer Science, 3(4), pp.1-10.

[12] Fuzhen Zhuang, Ping Luo, Hui Xiong, Qing He, Yuhong Xiong, and Zhongzhi Shi. Exploiting associations between word clusters and document classes for cross-domain text categorization. Statistical Analysis and Data Mining, 4(1):100-114, 2011.

[13] John Blitzer, Mark Dredze, and Fernando Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In Annual Meeting-Association For Computational Linguistics, volume 45, page 440, 2007.

[14] Bin Cao, Nathan Nan Liu, and Qiang Yang. Transfer learning for collective link prediction in multiple heterogenous domains. In Proceedings of the 27th International Conference on Machine learning, pages 159-166. Citeseer, 2010.

[15] Depin Chen, Jun Yan, Gang Wang, Yan Xiong, Weiguo Fan, and Zheng Chen. Transrank: A novel algorithm for transfer of rank learning. In Data Mining Workshops, 2008. ICDMW'08. IEEE International Conference on, pages 106-115. IEEE, 2008.

[16] Dataset is available at: https://www.kaggle.com/datasets/omarhatif/datasets-for-ear-detection-and-recognition

[17] Eyiokur, F.I., Yaman, D. and Ekenel, H.K., 2018. Domain adaptation for ear recognition using deep convolutional neural networks. iet Biometrics, 7(3), pp.199-206.

[18] Emeršič, Ž., Štepec, D., Štruc, V., Peer, P., George, A., Ahmad, A., Omar, E., Boult, T.E., Safdaii, R., Zhou, Y. and Zafeiriou, S., 2017, October. The unconstrained ear recognition challenge. In 2017 IEEE international joint conference on biometrics (IJCB) (pp. 715-724). IEEE.

[19] Alshazly, H., Linse, C., Barth, E. and Martinetz, T., 2019. Ensembles of deep learning models and transfer learning for ear recognition. Sensors, 19(19), p.4139.

[20] Shaha, M. and Pawar, M., 2018, March. Transfer learning for image classification. In 2018 second international conference on electronics, communication and aerospace technology (ICECA) (pp. 656-660). IEEE

[21] Eyiokur, F.I., Yaman, D. and Ekenel, H.K., 2018. Domain adaptation for ear recognition using deep convolutional neural networks. iet Biometrics, 7(3), pp.199-206.

[22] El-Naggar, S. and Bourlai, T., 2019, November. Evaluation of deep learning models for ear recognition against image distortions. In 2019 European Intelligence and Security Informatics Conference (EISIC) (pp. 85-93). IEEE.

[23] Zarachoff, M.M., Sheikh-Akbari, A. and Monekosso, D., 2021. Non-Decimated Wavelet Based Multi-Band Ear Recognition Using Principal Component Analysis. IEEE Access, 10, pp.3949-3961.

[24] Ahila Priyadharshini, R., Arivazhagan, S. and Arun, M., 2021. A deep learning approach for person identification using ear biometrics. Applied intelligence, 51(4), pp.2161-2172.

[25] Singh, S. and Suman, S., 2022, April. Transfer learning: A way for Ear Biometric Recognition. In 2022 IEEE 7th International conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.