J. Electrical Systems 20-3s (2024): 1397-1412

[1]Rahul Sawhney

[2]Shilpi Sharma

[3]Swapnita Srivastava

[4]Vipul Narayan

Ear Biometry: Protection Safeguarding Ear Acknowledgment Framework utilizing Transfer Learning in Industry 4.0

Abstract: - Human-Recognition using biometric features proposes a noble way of uniquely identifying individuals considering they do not require people to keep loads of passwords in mind to prove their identity. Biometrics have aided in people unable to render other's identity as well and have advanced over the years. Identification using ear biometric technique is considered to outrun other features since passive human involvement and ease-of-access are its strong set of attributes, not seen in any other biometric techniques. It has managed to identify criminals in the Crime Branch and has various helpful applications in the industry. Even in times of pandemic, it can serve its purpose in identifying individuals due to its feasibility. Models for Ear Image Identification have been proposed by various researchers over time utilizing Deep Learning and its Networks achieving high accuracy results, presenting faster and accurate identification models, boosting Ear Biometry as a secure Human Identification Tool. The learnings of this study on the AMI Ear Dataset and the OCEar dataset prove Ear Uniqueness of individuals demonstrating the identification of individuals, introducing passive identification into play, as well as computes on another dataset OCEar to study the similarity of both ears for identifying an individual.

Keywords: Ear Biometry, Ear Recognition, Convolutional Neural Networks, Human Ear, Transfer Learning.

I. INTRODUCTION

Over the years, Ear Biometry Research has seen rapid growth and continues to strive as a novel method of Human Recognition. The first instances of identifying Ear as a unique feature for Human Recognition can be seen mentioned by Iannareli [1] and to boost this statement, other studies backed by mathematical and neural models have been carried out. Ear Biometry proves its significance because of its easily feasible and accessible nature, passive involvement of human subjects and a plus point that it cannot be carbon copied by fraudsters easily.

Artificial Neural Networks and *Deep Learning* studies also come into play while adding more into this field. Deep Learning Networks have been proposed by various researchers in proving Ear Uniqueness and identifying individuals with the aid of Ears achieving great accuracy, aiding to the Human Identification process. In the papers proposed by Gonzalez, Alvarez and Mazorra [2], feature extraction and normalization techniques were used to identify unique regions and ear contours for identification, Ahila et al. [3] proposed Deep Network Model for identifying individuals in uncontrolled environments and Nejati et al. [4] succeeded in identifying twins from ear images, stating the fact that even twins can be identified using Ear images, thus proving its unique identification power. Artificial Neural Networks and *Deep Learning* (DL) methods have been studied for its applications in wide domains extending from pattern recognition and *Natural Language Processing* (NLP) to Computer Vision. Its implementation with Ear Biometry helps achieve robustness, efficiency and accuracy.

Since at least 1890, when French criminologist Alphonse Bertillon [1896] wrote4: "The ear, due to these countless little valleys and hills which furrow through it, is the most crucial factor from the point of identification," the ear's potential for this use has been recognized and supported. This organ is like the "intangible legacy of heredity and of the intra-uterine existence," remaining constant in shape from the moment of birth and impervious to the effects of environment and education throughout a person's whole lifetime. The ear's description and certain measurements were employed by Bertillon in his system of Bertilllonage, which was developed to track repeat offenders.

Iannarelli's approach was among the earliest ear recognition methods, created in 1949 [Iannarelli 1989]. As can be seen in Figure 5, this is a manually operated system involving a total of 12 measurements. Each ear picture is

[1,2]Department of Computer Science and Engineering, Amity University, Uttar Pradesh, India

[3] GL Bajaj Institute of Technology and Management

[4]Department of Computer Science and Engineering, Galgotias University, Uttar Pradesh, India

sawhney.rahulofficial@outlook.com , ssharma22@amity.edu,swapnitasrivastava@gmail.com,vipulupsainian2470@gmail.com

positioned such that the lower tip of a standardized vertical guide on the development easel reaches the upper flesh line of the cocha region and the higher tip contacts the contour of the antitragus. After that, the helix's crus are located and used as the pivot. From that point, we may draw lines that are vertical, horizontal, diagonal, and anti-diagonal to meet the pinna's internal and exterior curves at various locations. These points of intersection serve as the basis for the 12 ear measurements.

The first completely automated method for ear recognition was described by Moreno et al. [1999]. They employed many features and averaged the classification accuracy of numerous neural networks. Their feature vector had macro-features retrieved by a compression network in addition to micro-features such as the outer ear points, ear shape, and wrinkles. Two different picture sets were taken to put the system to the test. There was a total of 168 photographs in the initial set, and each of the 28 themes was shown in 6 pictures. The second group consisted of 20 photos representing 20 unique people. This technique was further expanded upon by Mu et al. [2004]. The exterior and inner ear structures were combined to form a single vector that they used to represent the ear's feature vector. The neural network was then used for categorization. This approach is like the system proposed by Iannarelli [Iannarelli, 1989], but with fewer moving parts. To reduce the mean square error between the probe and gallery photographs, Yuizono et al. [2002] approached the ear image identification problem as a standard search optimization issue, solving it using a Genetic Algorithm (GA). A total of 660 photos representing 110 unique individuals were compiled. They had a success percentage of 99.99 percent. Ear recognition research follows the same pattern as that of other biometric traits: it is guided by the availability of datasets for algorithm assessment and performance analysis. As a result, we first go through the many datasets that have been compiled by different research teams with the purpose of gauging the feasibility of ear biometrics.

There are many applications and problem recognitions that have given rise to new and improved frameworks in various fields. Some examples include:

1. Data science and machine learning: There has been a proliferation of data in recent years, leading to the need for more advanced methods of analysis and prediction. This has led to the development of new frameworks such as deep learning, which has had significant success in tasks such as image and speech recognition.

2. Software engineering: As software systems have become more complex, there has been a need for better ways to design and build them. This has led to the development of frameworks such as agile methodologies, which emphasize flexibility and adaptability in the software development process.

3. Environmental sustainability: As the impacts of human activity on the environment have become more evident, there has been a push for new frameworks that address environmental challenges. For example, the circular economy framework seeks to minimize waste and maximize resource efficiency, while the sustainable development goals provide a framework for addressing global environmental and social challenges.

4. Social justice: There have been various efforts to address social and economic inequalities, leading to the development of frameworks such as intersectionality, which recognizes the interconnected nature of social categorizations such as race, class, and gender and their roles in systemic discrimination and disadvantage.

5. Education: The rapid pace of technological and societal change has led to a need for more flexible and adaptable education systems. This has led to the development of frameworks such as competency-based education, which focuses on students demonstrating their knowledge and skills rather than simply completing a set of courses.

These are just a few examples of the many applications and problem recognitions that have given rise to new and improved frameworks. In each case, the development of these frameworks has been driven by the need to address specific challenges or opportunities in a more effective and efficient manner. As new problems and opportunities arise, it is likely that new frameworks will continue to be developed to address them.

There are several techniques that have been used to identify unique regions and ear contours for identification purposes. Some of these techniques include:

1. Feature extraction: This involves extracting specific features from an ear image that are unique to an individual and can be used to identify them. These features could include the shape of the ear, the size and position of the lobes, the presence of scars or other markings, and the overall symmetry of the ear.

2. Ear recognition algorithms: There are several algorithms that have been developed specifically for ear recognition, such as the convolutional neural network (CNN) and the local binary pattern (LBP) algorithm. These algorithms analyse the ear image and compare it to a database of known ear images to identify a match.

3. 3D ear modelling: Another technique involves creating a 3D model of an ear based on a series of images taken from different angles. This allows for a more detailed and accurate representation of the ear's shape and features, which can be used for identification purposes.

4. Ear print analysis: Some researchers have focused on the unique patterns of the ear's ridges and valleys, known as an ear print, as a means of identification. By analysing the ear print, it is possible to create a unique identifier for an individual.

5. Fusion of multiple modalities: In some cases, it may be useful to combine multiple techniques for improved accuracy. For example, combining ear recognition algorithms with 3D ear modelling or ear print analysis can improve the overall performance of the identification system.

Overall, the choice of technique for identifying unique regions and ear contours will depend on the specific requirements of the application and the resources available. Some techniques may be more accurate or efficient in certain situations, while others may be more practical or cost-effective. Researchers are continually developing and improving upon these techniques to enhance the accuracy and reliability of ear recognition systems.

The proper region of the ear that needs to be extracted using cropping techniques will depend on the specific application and the requirements of the system. In general, it is important to include as much of the ear as possible in the cropped image, while still maintaining the integrity of the ear's features.

There are several factors to consider when determining the proper region of the ear to be cropped:

1. Resolution: The resolution of the cropped image should be sufficient to allow for the accurate identification of the ear's features. This may require cropping a larger region of the ear to capture more detail.

2. Aspect ratio: The aspect ratio of the cropped image should be kept in mind to ensure that the ear's features are not distorted. For example, if the ear is wider than it is tall, the cropped image should reflect this.

3. Positioning: The ear should be properly positioned in the cropped image, with the features of the ear clearly visible. This may require adjusting the position of the ear or the camera to ensure that the desired features are captured.

Cropping techniques involve selecting a specific region of an image and removing the rest. There are several ways to perform cropping, including:

1. Manual cropping: This involves physically selecting the region of the image to be cropped using a tool such as a rectangle or ellipse. This method is typically used for smaller images or for making minor adjustments to the crop.

2. Automatic cropping: This involves using algorithms or software to automatically identify the region of the image to be cropped. This can be useful for large images or for cropping multiple images in a batch.

3. Aspect ratio cropping: This involves cropping the image to a specific aspect ratio, such as 4:3 or 16:9. This can be useful for ensuring that the image will fit properly in a specific size or shape.

Overall, the choice of cropping technique will depend on the specific requirements of the application and the resources available. Some techniques may be more accurate or efficient in certain situations, while others may be more practical or cost-effective. It is important to carefully consider the proper region of the ear to be cropped and the best cropping technique to use to ensure the accuracy and reliability of the ear recognition system.

Feature extraction and normalization are techniques that are often used in ear recognition systems to identify unique regions of the ear. Feature extraction involves extracting specific features from an ear image that are unique to an individual and can be used to identify them. These features could include the shape of the ear, the size and position of the lobes, the presence of scars or other markings, and the overall symmetry of the ear. There are several methods for extracting features from an ear image, including:

1. Edge detection: This involves identifying the edges of the ear and creating a map of the ear's contours. This can be useful for identifying specific features such as the lobe or the outline of the ear.

2. Texture analysis: This involves analysing the patterns and textures of the ear's surface, such as the ridges and valleys of the ear's skin. This can be useful for identifying unique patterns that are specific to an individual.

3. Shape analysis: This involves analysing the overall shape of the ear and identifying specific features such as the curvature of the ear or the presence of certain points or angles.

Normalization is the process of standardizing the ear image so that it can be compared to other ear images in a consistent manner. Normalization may be necessary to account for variations in the size, orientation, or lighting of the ear images. Some methods for normalizing ear images include:

1. Resizing: This involves adjusting the size of the ear image so that it is the same size as other ear images in the database. This can be useful for ensuring that the ear images are comparable.

2. Rotation: This involves adjusting the orientation of the ear image so that it is facing the same direction as other ear images in the database. This can be useful for ensuring that the ear images are comparable.

3. Lighting correction: This involves adjusting the lighting of the ear image to account for variations in lighting conditions. This can be useful for ensuring that the ear images are comparable [25][26].

Overall, feature extraction and normalization are important techniques for identifying unique regions of the ear and ensuring that ear images are comparable. These techniques can be used in combination with ear recognition algorithms and other methods to improve the accuracy and reliability of ear recognition systems [27][28].

Convolutional neural networks (CNNs) are a type of artificial neural network specifically designed for image recognition tasks. They are inspired by the way the human brain processes visual information and can be used for tasks such as image classification, object detection, and image generation. The mirror neurons in the human brain are a type of neuron that activates when an individual observes or performs an action, as well as when they imagine performing the action. These neurons are thought to play a role in empathy and understanding the actions and intentions of others. In a CNN, the neurons are organized into layers, with each layer responsible for processing a specific aspect of the input image. The first layer, known as the input layer, receives the raw image data. The subsequent layers, known as hidden layers, apply a series of mathematical transformations to the image data to extract features such as edges, shapes, and patterns. The final layer, known as the output layer, uses the extracted features to make a prediction or decision based on the input image. One key feature of CNNs is their use of convolutional layers, which apply a set of filters to the image data to extract features. These filters are responsible for detecting specific patterns or features in the image, such as edges or shapes. The filters are learned during the training process, allowing the CNN to adapt to the specific characteristics of the input images. Another key feature of CNNs is their use of pooling layers, which down sample the image data to reduce the computational burden and improve the generalization ability of the network. Pooling layers apply a function, such as max pooling or average pooling, to a small region of the image to reduce its dimensionality. Overall, CNNs are a powerful tool for image recognition tasks, with their ability to learn and adapt to specific image characteristics and extract relevant features. They are inspired by the way the human brain processes visual information and can be used to perform a wide range of tasks.

There are several methods for extracting features from images, including:

1. Edge detection: This involves identifying the edges of objects in an image and creating a map of the image's contours. This can be useful for identifying specific features such as the outline of an object or the boundaries between different regions.

2. Texture analysis: This involves analysing the patterns and textures of the image, such as the ridges and valleys of an object's surface. This can be useful for identifying unique patterns that are specific to an object.

3. Shape analysis: This involves analysing the overall shape of an object and identifying specific features such as the curvature or the presence of certain points or angles [29][30].

Deep convolutional models (DCMs) are a type of artificial neural network specifically designed for image recognition tasks. They are inspired by the way the human brain processes visual information and can be used for tasks such as image classification, object detection, and image generation. DCMs are composed of multiple layers of artificial neurons, with each layer responsible for processing a specific aspect of the input image. The first layer, known as the input layer, receives the raw image data. The subsequent layers, known as hidden layers, apply a series of mathematical transformations to the image data to extract features such as edges, shapes, and patterns. The final layer, known as the output layer, uses the extracted features to make a prediction or decision based on the input image. One key feature of DCMs is their use of convolutional layers, which apply a set of filters to the image data to extract features. These filters are responsible for detecting specific patterns or features in the image, such as edges or shapes. The filters are learned during the training process, allowing the DCM to adapt to the specific characteristics of the input images. Another key feature of DCMs is their ability to learn deep representations of the image data, meaning that they can extract high-level features from the data by processing it through multiple layers. This allows DCMs to perform well on complex image recognition tasks [31].

General adversarial networks (GANs) are a type of artificial neural network that consists of two separate networks: a generator network and a discriminator network. The generator network is responsible for generating new, synthetic data that is like the real data, while the discriminator network is responsible for distinguishing between real and synthetic data. The generator and discriminator networks are trained together, with the generator attempting to produce synthetic data that is indistinguishable from the real data, and the discriminator attempting to accurately distinguish between real and synthetic data. This process can be thought of as a "game" between the two networks, with the generator trying to "fool" the discriminator and the discriminator trying to correctly identify the synthetic data. GANs have been used for a variety of tasks, including image generation, style transfer, and data augmentation. They can produce high-quality synthetic data and have been used in a variety of applications, such as creating realistic images of objects or people.

Ear biometry is the study of the shape and size of the human ear as a means of identification. It has gained significant attention in recent years because of its potential as a biometric identifier, as the ear is a unique and stable characteristic that can be easily captured and measured. Ear biometry is a reliable method of identification because the ear is a complex and highly detailed structure with many unique features. The ear's shape, size, and other characteristics are largely determined by genetics and are not easily altered, making it a suitable biometric identifier. One of the main advantages of ear biometry is its accessibility. The ear is easily visible and can be easily captured using a standard camera or smartphone, making it convenient to use as a biometric identifier. This contrasts with other biometric identifiers such as fingerprints or iris scans, which may require specialized equipment or proximity to the subject.

Ear biometry is also relatively non-invasive and non-intrusive, making it more acceptable to individuals compared to other biometric identifiers that may require physical contact or the collection of sensitive personal information. Ear biometry has several potential applications, including security, identification, and authentication. It can be used to verify the identity of individuals in a variety of settings, such as border control, financial transactions, and access control. Ear biometry has the potential to improve the accuracy and reliability of identification systems, as it is less prone to errors or false matches compared to other biometric identifiers. It is also less susceptible to tampering or counterfeiting, as the ear is not easily altered. Overall, ear biometry is a promising biometric identifier due to its accessibility, non-invasiveness, and high degree of uniqueness. Its potential applications and advantages make it a valuable tool for identification and authentication in a variety of settings.

There are several types of neural networks that involve competition between multiple networks to refine and improve their performance. One such type of neural network is the adversarial network. Adversarial networks consist of two separate networks: a generator network and a discriminator network. The generator network is responsible for generating new, synthetic data that is like the real data, while the discriminator network is responsible for distinguishing between real and synthetic data. The generator and discriminator networks are trained together, with the generator attempting to produce synthetic data that is indistinguishable from the real data, and the discriminator attempting to accurately distinguish between real and synthetic data. This process can be thought of as a "game" between the two networks, with the generator trying to "fool" the discriminator and the discriminator trying to correctly identify the synthetic data. The competition between the generator and discriminator networks serves to refine and improve the performance of both networks. As the generator produces synthetic data, the discriminator must learn to accurately distinguish between real and synthetic data. This in turn improves the ability of the generator to produce synthetic data that is indistinguishable from the real data.

Another type of neural network that involves competition between multiple networks is the self-organizing map (SOM). SOMs are used for clustering and visualization of high-dimensional data. There are several types of neural networks that involve competing or cooperative processes to refine their performance and reduce losses. Some examples include:

1. Generative adversarial networks (GANs): These are a type of neural network that consists of two separate networks: a generator network and a discriminator network. The generator network is responsible for generating synthetic data that is like real data, while the discriminator network is responsible for distinguishing between real and synthetic data. The two networks are trained together, with the generator attempting to produce synthetic data that is indistinguishable from real data and the discriminator trying to correctly identify the synthetic data. The process can be thought of as a "game" between the two networks, with the generator trying to "fool" the discriminator and the discriminator trying to accurately identify the synthetic data.

2. Co-training: This is a type of semi-supervised learning where two separate classifiers are trained on different views of the data, such as different feature sets or sources. The classifiers are then used to label the remaining unlabelled data, and the process is repeated until the classifiers reach a satisfactory level of performance.

3. Ensemble learning: This involves training multiple classifiers and combining their predictions to improve the overall accuracy of the model. The classifiers can be trained on different versions.

The model proposed in this paper focuses on two objectives: (i) Proving Ear Uniqueness using Deep Convolutional Networks; (ii) Stating the similarity of both ears in identifying an individual.

The following sections move along the following stated sections: Section 2 gives discussion on the Human Ear and provides Literature Review; Section 3 discusses the Deep Convolutional models used in this paper; Section 4 gives information on Dataset used; Section 5 explains the methodology; Section 6 displays experiment results; Section 7 gives the conclusions giving way for further research on the subject. Lastly, Section 8 focuses on Discussions.

II. THE HUMAN EAR

The Human Ear located symmetrically on the centres over the temporal lobe of the head allows humans to receive auditory signals and takes the command of the position sense and balance. The Ear is divided in three study areas: Outer, Middle and Inner respectively. The readings of this paper is focused on the Outer Ear called the *Pinna* (Fig. 1). It proves to be a great identification part since it remains the same throughout life, although the lobule elongates downwards which does not propose any affect towards the structure much and also the Ear shape does not change as do the facial expressions. Although some minor degree of unparallel uniformity may be observed in both the ears of an individual [5], distinguishing between the individuals is a possibility regardless of any external factors.

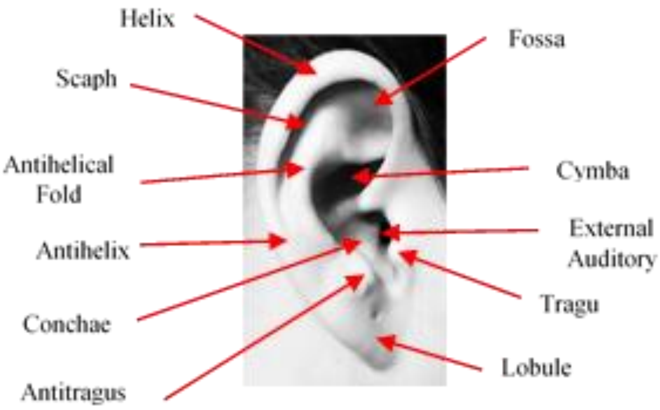

Figure. 1. Anatomy of the Outer Ear

A. Motivation and Literature Review

The Human Ear has been widely experimented for uniqueness using geometric and feature extraction methods along with Deep Convolutional Models and General Adversarial Networks over time, proving noble ways of Human-Identification, and improving models over the ages, furthermore being able to generate a person's face with a mere image of one's ear [6]. As early as in 2000, M. Burge and W. Burger [7] proposed a graph technique for authentication that draws out the curvings of the ear and matches features based on those extracted curves. In the proceedings by Hoogstrate, A. J., Van Den Heuvel, H., & Huyben, E. [8], they worked on identification of individuals based on surveillance camera images, while in papers proposed by M. Choraś [9], J. D. Bustard and M. S. Nixon [10], Pflug, Anika and C. Busch. [11] and K. Annapurani, M.A.K. Sadiq, and C. Malathy [12], the focus for classification was built on geometric and feature extraction processes. The issue of Ear Segmentation from images for identification purpose proposed by E. H. Said, A. Abaza and H. Ammar [13], devised a mathematical model which achieved an accuracy of over 90% for distinguishing ear samples from facial images. In another paper proposed by Li Yuan, Zhichun Mu [14], they combined Gabor feature filter and KFDA (Kernel Fisher Discriminant Analysis) methods for Ear Recognition and Authentication on 2D images.

With the advancement of DL models and Deep Neural Networks (DNN) in recent years, researchers have integrated these models into improving Ear Classification architectures. Among Artificial Neural Network architectures, *Convolutional Neural Networks* (CNN) emerges in the top for image processing and recognition. The Convolutional Neural Networks mirror the functioning of neurons in the human brain, and the neurons that make up the layers of a CNN model are mathematical functions to calculate weighted sum of multiple inputs and output an activation value highlighting the relevant features of the input image. Based on these output values of the last convolutional layer, the final classification layer gives a probability of the image being from one of the classes.

In 2018, Cintas, C. et al. [15], proposed a model that exploited the geometric morphometry of the ear and CNN architecture to work on the features of the ear as plotted points mirroring the individual's ear shape with which the model was able to correctly form the geometric morphometry of the ears and additionally it also performed well on the recognition process with a precision value of 95% and a recall value of 90%. Eyiokur, F.I., Yaman, D., & Ekenel, H.K. [16] also performed Ear Recognition with *Deep Convolutional Neural Networks* (DCNN) and *Transfer Learning* using pretrained models of AlexNet, VGG-16 and GoogLeNet.

In the discipline of Ear Biometry and Ear Recognition, successful implementation of GANs (*General Adversarial Networks*) can also be seen since 2020. General Adversarial Networks are deep generative models that consist of two networks/models within, namely, the *Generator* and the *Discriminator* networks. These two networks compete amongst each other to refine the other and reduce losses. They find applications in image generation, text-to-image conversion (StackGAN), MRI image reconstruction, Object Transfiguration (CycleGAN), etc. In the realm of Ear Recognition, GANs can be seen in the analysis paper proposed by Khaldi, Y., Benzaoui, A [17], where DCGAN (Deep Convolutional GAN) model and CNN model have been used to recognise ears from grayscale pictures.

Ear biometry is a method used to measure and analyze the size and shape of the ear for the purpose of identification or for creating ear prosthetics. It is often used in forensic science and in the field of plastic and reconstructive surgery.

There are several methods of ear biometry, including manual measurement with calipers, photogrammetry, and 3D scanning. Zhu, Q.; Mu, Z [23] Manual measurement with calipers is the most traditional method and involves taking linear and angular measurements of the ear using a physical tool. Photogrammetry involves taking multiple photographs of the ear from different angles and using software to measure the ear based on the photographs. 3D scanning involves using a specialized device to create a digital model of the ear, which can then be measured using computer software.

One of the main benefits of ear biometry is that the ear is a relatively stable and unique feature of the human body, making it a useful tool for identification purposes. Ramos-Cooper, S [24] In forensic science, ear biometry can be used to help identify individuals from partial remains or from photographs. In plastic and reconstructive surgery, ear biometry can be used to create customized ear prosthetics or to plan ear reshaping surgery.

There are some limitations to ear biometry, however. The accuracy of the method can be affected by factors such as the quality of the images or scans used, the experience of the person taking the measurements, and the inherent variability in ear shape. In addition, ear biometry is not always reliable for identifying individuals of different racial or ethnic backgrounds, as there can be greater variation in ear shape among these groups.

Overall, ear biometry is a useful tool for identification and surgical planning, but it is important to consider its limitations and to use it in conjunction with other methods as needed

It is quite common for people to have some degree of unparallel uniformity in the shape and size of their ears. This is because the human body is not perfectly symmetrical and there can be minor variations in the way different parts of the body develop. While the ears may not be the same size or shape, they are usually relatively close in terms of their overall appearance. One of the reasons for the observed unparallel uniformity in the ears is genetics. Our genetic makeup plays a significant role in determining the physical characteristics of our body, including the shape and size of our ears. Some people may inherit certain traits that cause their ears to be slightly different in size or shape from one another.

Another reason for the unparallel uniformity in the ears may be due to environmental factors. For example, exposure to certain substances or conditions during development, such as certain medications or toxins, can affect the way the ears develop. Additionally, physical trauma or injury to the ears during childhood or later in life can also cause them to be uneven. Despite the observed unparallel uniformity in the ears, they typically function normally and do not cause any significant problems. In some cases, people may choose to undergo cosmetic surgery to alter the appearance of their ears if they are particularly asymmetrical or if they feel self-conscious about the way they look. However, such procedures are generally considered elective and are not medically necessary. Overall, the observed unparallel uniformity in the ears is a normal and common variation in human anatomy. While they may not be the same size or shape, the ears usually perform their essential functions without any issues.

Thus, over the years, the field of research on Ear Biometry and its expanding applications and problem recognitions have certainly given rise to new and improved frameworks and with the rising advancement of DNNs and architectures, problems that were of concern are being solved.

III. DATASET USED

The Datasets employed for the verification of this study have been collected from two sources:

(i) AMI Ear Dataset.

(ii) OCEar Dataset.

(i) AMI Ear Dataset: This dataset contains 700 images; each of the 100 individuals has 7 images ranging from the age of 19 to 65 years. Out of those 7 images, for the right ear we have 6 images and 1image for the left. The 6 images of the right ear are taken from right, up, down, left, front and zoomed angles and all images have a resolution of 492×702 pixels. Some of the AMI Ear dataset images are shown in Fig. 2.

(ii) OCEar Dataset: The OCEar Dataset makes up 114 images collected from the people containing Left and Right Images of 57 individuals aged between 6 and 55 years. The image resolutions range over a wide range from 287×510 to 2440×3303 pixels. The images are unconstrained unlike the previous Ear Dataset as the images vary in lighting and angles. The OCEar dataset images are shown in Fig. 3.

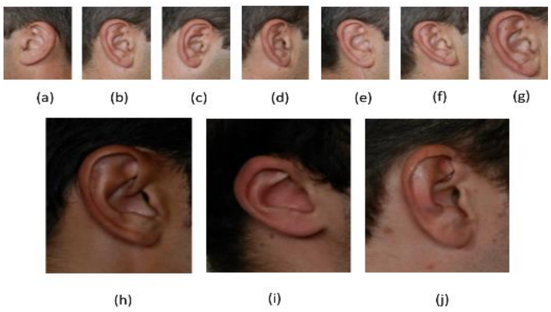

Figure. 2. Sample images of AMI Ear Dataset; (a)back view; (b)front view; (c)down view; (d)left ear; (e)right ear; (f)up view; (g)zoom view; (h)-(j) Random sample images

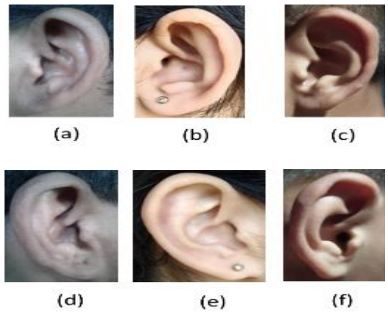

Figure 3. Sample images of OCEar Dataset; (a)-(c) Left Ears; (d)-(f) Right Ears

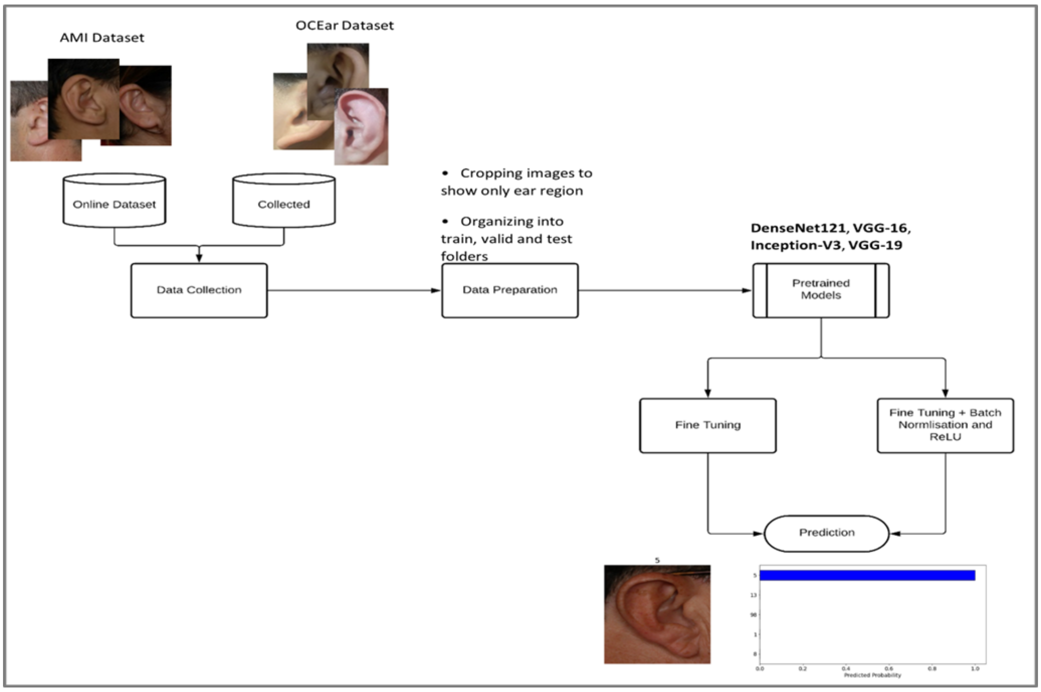

Figure 4. Approach of Paper for Ear Biometry Image Classification

IV. DEEP CONVOLUTIONAL NEURAL NETWORKS

Deep Convolutional Neural Networks have seen profound significance and applicability in the present years over Computer Vision. Unlike the CNN architecture with 5 to 10 feature extracting layers, the *Deep Convolutional Neural Networks* (DCNN) models have layers extending from 50 to 100 and even 1000 layers deep. First, these extract the low-level features and as the layers go deep, the higher-level features get extracted. With Image Classification problems, Deep Neural Models have proved useful and in this paper, discussions on DenseNet121, VGG-16, VGG-19 and InceptionV3 have been carried out as they have been applied for Transfer Learning.

A. DenseNet121

The DenseNet121 is one of the Dense Convolutional Network (DenseNet) [18] models, and as its name suggests, it has 5+(6+12+24+16)*2=121 layers, where 5 comes from (conv, pooling) + 3 transition layers + 1 classification layer, and the dense blocks i.e. (6+12+24+16) have been multiplied by 2 as each dense block has 2 layers, namely 1×1 conv and 3×3 conv. The DenseNets are a logical extension of ResNet [19] model. The trouble that DenseNet solves is that in the CNNs, as the model architecture goes deeper, information gets lost before reaching the other end, so in DenseNet, every layer is connected to every other layer in a feed-forward fashion.

B. VGGNet

The VGGNets [20] (named after Visual Geometry Group from Oxford) are Convolutional Neural Network architectures that take an input image of size 224×224 and passes them through a stack of the *convolutional blocks* to final fully connected layers that consist of 4096 channels to predict 1000 labels. It has two variants, namely, the VGG-16 [20] and the VGG-19 [20] models. VGG-16 has 13 convolutional layers, with 5 pooling layers in between and 3 fully connected (FC) layers, hence named to be VGG-16. As for the VGG-19 model, it has 19 convolutional layers with the same 5 pooling layers as is present in VGG-16 and 3 fully connected layers. The VGG models secured the best performance models in the ILSVRC-2014 challenge.

C. Inception-V3

The Inception-V3 [21] model is one of the networks from the Inception family that is 48 layers deep and accepts an input image of minimum size value upto 299×299. The Inception networks focus on sparse connected networks throughout the deep architecture, thus increasing the depth and width of the network. Considerably, it is more efficient than VGGNet and is considered as one of the best models to be implemented on devices with low processing units.

V. METHODOLOGY

In this work, four pretrained models are employed and compared for the classification of Ear Images and two datasets, the AMI and OCEar dataset were utilised for the process. The paper has two objectives, to prove Ear Uniqueness and Human Recognition possibility with four models of Image Classification, and secondly, to establish if both ears are similar, to prove that both the ears can be taken into account for identifying an individual. Fig. 4 gives the methodology followed in this paper.

A. Data Collection

For carrying out the experiments, AMI and OCEar datasets were collected from online sources. The OCEar dataset has been formed by images collected online from individuals aged between 6 and 55 years of age and has been clicked by the individuals themselves, thus have varying angles of capture and lighting.

B. Data Preparation

For this stage, which is the *data preparation* stage, the quantity of the datasets were increased for better training using data augmentation and divided into train, validation and test folders. The AMI dataset originally contained 700 images which was increased to 2000 images and the OCEar dataset contained 114 images which was increased to double its size i.e. 228 images. The OCEar dataset was processed for refinement over region of interest while the AMI dataset did not need any refinements. As shown in Fig. 5, the OCEar dataset had images clicked by the individuals at different distance measurements from the camera to the ear, so the proper region of ear needed to be extracted using cropping technique for the model to be used for the classification process.

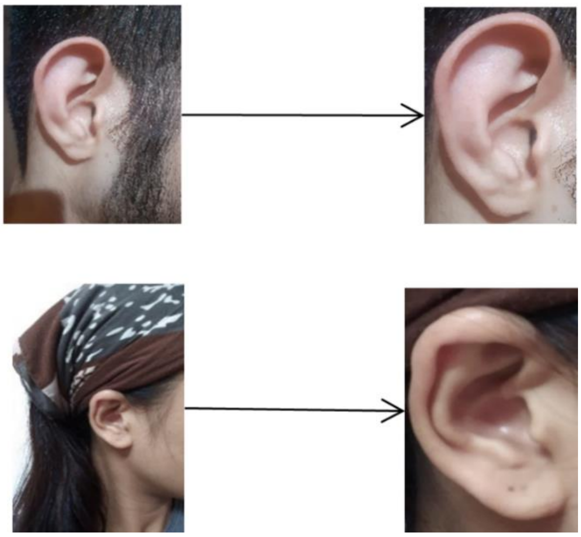

Figure. 5. OCEar image data preparation

The AMI Dataset images were divided into 5 images per training folder, 4 for the validation folder and 2 images per testing folder, and in the OCEar dataset, each training folder has 1 image, 2 images per validation folder and 1 image per testing folder.

C. Proposed Framework

Following the data preparation stage is the transfer learning stage where the 4 pretrained models are trained upon the datasets. In this stage, each dataset is trained twice on each model, once without fine tunings and a second run with fine tuning along with Batch Normalisation (BN) and ReLU layer additions. Considering, that the models were pretained on ImageNet dataset [22], Ear datasets possess different set of features that require a much deeper ability to be classified as they possess similar parts though different in sizes and lengths. Thus the pretrained models were fine tuned and layers of Batch Normalisation and ReLU were added in some models.

The DenseNet121 pretrained model did not need the addition of any extra layers except for the tweaks in hyperparameters and so did not require to be run twice for both AMI and OCEar datasets. However, for Inception-V3, VGG-16 and VGG-19 models, along with the hyperparameter tweakings and fine tunings, additional layers were implemented to increase model precision and accuracy.

For VGG-16 and VGG-19 with Fine Tunings, the last fully connected (FC) layer and output layer were replaced by four fully connected layers, three ReLU actvation filters and three dropout layers followed by a LogSoftmax output layer. In the second experimental run with additional layers, Batch Normalisation layers are placed before every MaxPool Layer in both the models along with the fully connected layers, ReLU activations and a LogSoftmax output layer. For Inception-V3 with Fine tunings, the fully connected layer is replaced by four fully-connected layers each of the first three followed by a ReLU activation filter and a Dropout Layer. The last fully-connected layer is followed by a LogSoftmax output layer. And in the second experimental run with additional layers, ReLU activation functions are placed before the Batch Normalisation (BatchNorm2d) Layers of Mixed_7a:InceptionD blocks, Mixed_7b and Mixed_7c:InceptionE blocks. These additions are for the AMI Ear Dataset, while for OCEar dataset only the Mixed_7c:InceptionE blocks need addition of the ReLU activations.

VI. EXPERIMENT RESULTS

For the results of the models on the datasets the following performance metrics are evaluated:

- Precision
- Recall
- F1 Score
- Confusion Matrix

In the experiments carried out for the AMI Ear Dataset, DenseNet121 surpasses all the other models even without addition of extra layers and with lesser parameter trainings as compared to other models. Also, we can witness the increase in accuracies of all the other pretrained models with the extra additional layers of Batch Normalisation and ReLU. The low performances of the metrics on fine tuning can be due to the learned weights on ImageNet and on adding layers and training some layers with the ear dataset images, the results generated show gradual increase in the prediction results. Table 1 and Table 2 give the number of parameters and performance tabulation for all the models on the AMI Ear Dataset. For the results of OCEar dataset given in Table 3 and Table 4, we can see that not great results were generated by the experiment and the reasons can be many. Firstly, the dataset has train and test folders with just one images for each of the classes, and that blends in difficulty for the models to conduct better training on the OCEar dataset. Also, the images are taken under no guidance with ear images of the same individual being clicked from different angles as seen above in the dataset samples given in Figure 3. However, we can also see that the models could determine an individual's right from the left image given as training data even with the absence of sufficient images, from which we can understand that both the ears of individuals are recognizable as one and thus pose no threat to Ear Biometry. Hence identification of an individual using both of their ears is a possibility and is a statement verified.

For the Final stage, the best performing model DenseNet121 is made to predict the results of the provided pictures given in Fig 6. The DenseNet121 model predicts the individuals with the given ear images of both the datasets and this concludes the statement that Neural Networks can be used for Ear Biometry purpose and used for identifying individuals.

TABLE-1: Total and Trainable Parameters for models on AMI Ear Dataset

Model Tuning Type	Model	Total Parameters	Trainable Paramters
Fine Tuning	DenseNet121	7,529,956	576,100
	VGG-19	150,635,172	11,064,932
	Inception-V3	27,813,804	2,701,540
	VGG-16	145,325,476	11,064,932
Fine Tuning + Batch Normalisation and ReLU	DenseNet121	7,529,956	576,100
	VGG-19	150,638,116	11,067,876
	Inception-V3	27,813,804	15,505,636
	VGG-16	145,328,420	11,067,876

TABLE-2: Classification process results on AMI Ear Dataset

Model Tuning Type	Model	Precision	Recall	F1 Score
Fine Tuning	DenseNet121	**95.47**	**95.00**	**94.34**
	VGG-19	42.53	39.00	36.80
	Inception-V3	48.79	51.5	46.85
	VGG-16	64.39	57.99	56.70
Fine Tuning + Batch Normalisation and ReLU	DenseNet121	**95.47**	**95.00**	**94.34**
	VGG-19	81.96	78.50	77.05
	Inception-V3	87.07	85.50	83.80
	VGG-16	81.20	81.50	78.84

TABLE-3: Total and Trainable Parameters for models on OCEar Dataset

Model Tuning Type	Model	Total Parameters	Trainable Paramters
Fine Tuning	DenseNet121	7,551,673	597,817
	VGG-19	150,613,113	11,042,873
	Inception-V3	27,808,257	2,695,993
	VGG-16	145,303,417	11,042,873
Fine Tuning + Batch Normalisation and ReLU	DenseNet121	7,551,673	597,817
	VGG-19	150,616,057	11,045,817
	Inception-V3	27,808,257	8,766,265
	VGG-16	145,306,361	11,045,817

TABLE-4: Classification process results on OCEar Dataset

Model Tuning Type	Model	Precision	Recall	F1 Score
Fine Tuning	DenseNet121	**52.34**	**64.91**	**56.43**
	VGG-19	23.09	31.58	25.44
	Inception-V3	23.59	31.58	25.47
	VGG-16	29.49	36.84	31.14
Fine Tuning + Batch Normalisation and ReLU	DenseNet121	**52.34**	**64.91**	**56.43**
	VGG-19	34.79	43.86	37.43
	Inception-V3	30.70	38.59	32.66
	VGG-16	44.88	54.38	47.78

The confusion matrices of the models can better visualize the results of the classification models used in this paper.

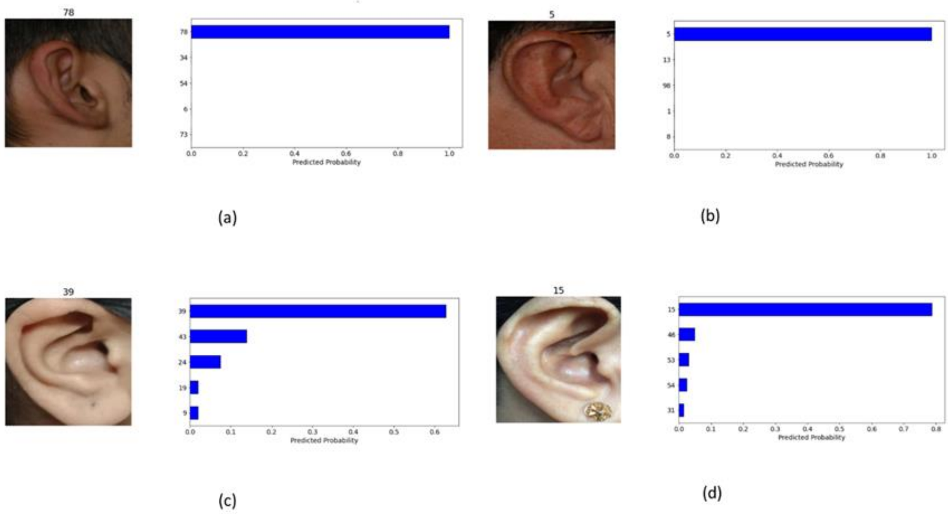

Fig. 6. DenseNet121 Prediction results on AMI Dataset and OCEar dataset; (a)-(b) AMI predictions; (c)-(d)

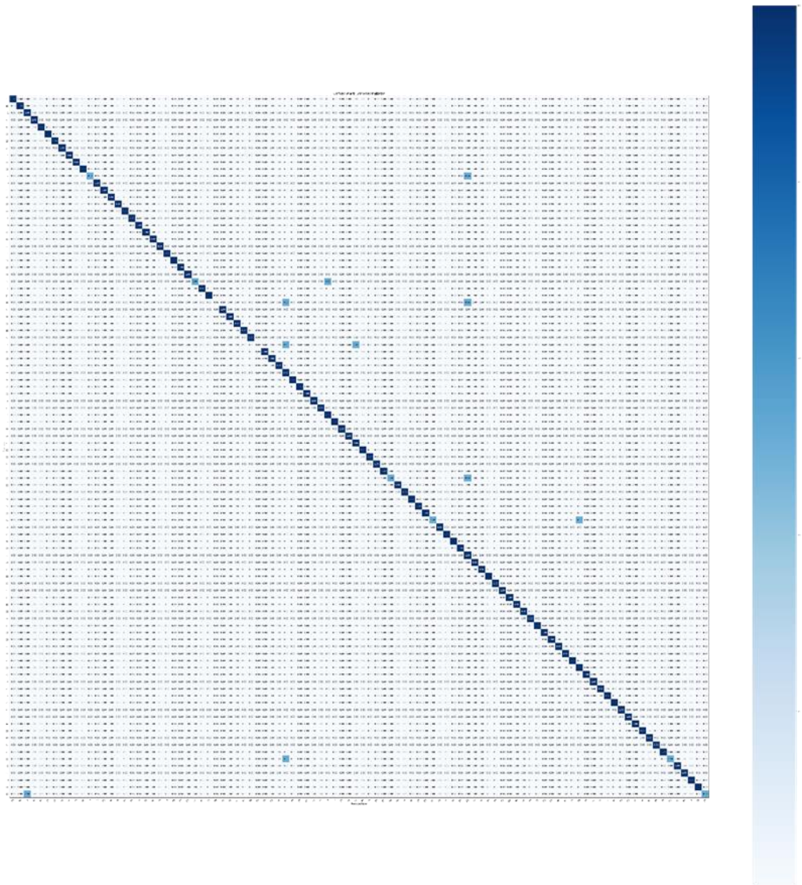

Figure. 7.a. DenseNet121 Confusion Matrix on AMI with Fine tuning

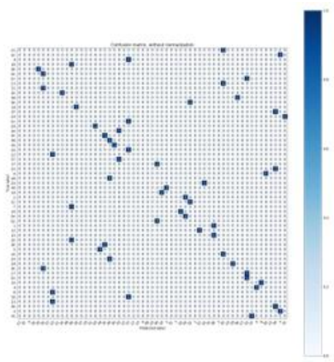

Figure 8.b. VGG-19 Confusion Matrix on OCEar with Fine Tuning

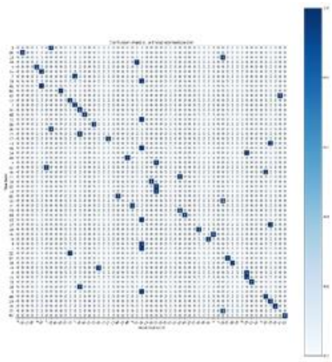

Figure. 8.c. VGG-19 Confusion Matrix on OCEar with Fine Tuning + Batch Normalisation and ReLU

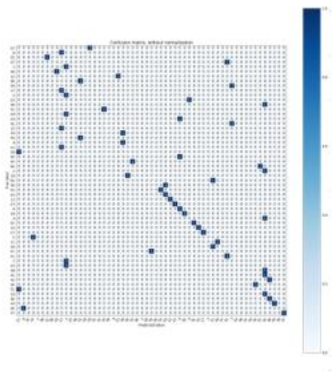

Figure. 8.d. Inception-V3 Confusion Matrix on OCEar with Fine Tuning

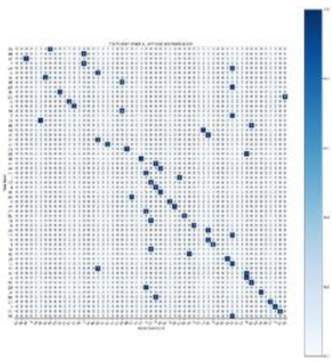

Figure. 8.e. Inception-V3 Confusion Matrix on OCEar with Fine Tuning + Batch Normalisation and ReLU

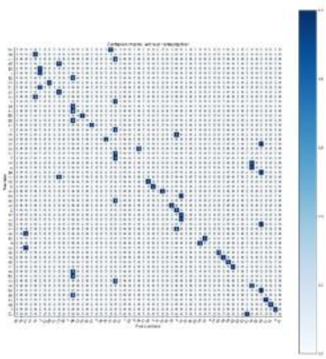

Figure 8.f.VGG-16 Confusion Matrix on OCEar with Fine Tuning

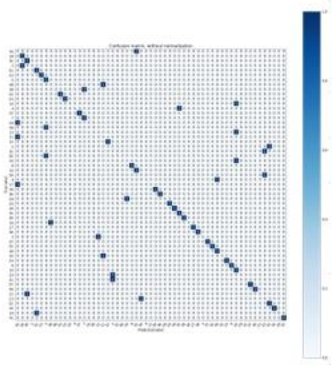

Figure. 8.g. VGG-16 Confusion Matrix on OCEar with Fine Tuning + Batch Normalisation and ReLU

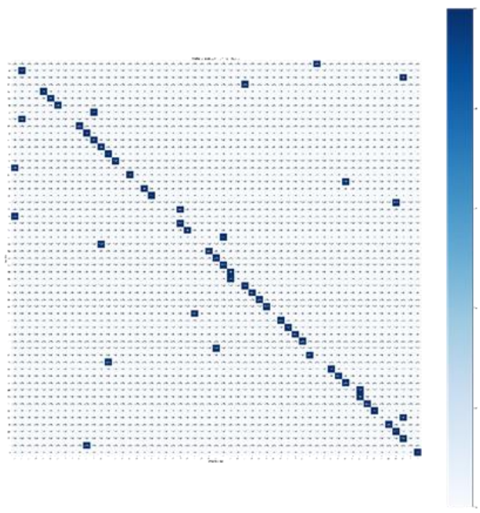

Figure. 8.a. DenseNet121 Confusion Matrix on OCEar with Fine tuning

OCEar predictionsFrom Fig 7.a. to Fig 7.g., confusion matrices on AMI dataset for all models have been displayed and Fig 8.a. to Fig 8.g. gives the confusion matrices on the OCEar datase for all models. Confusion Matrices as performance measurer perform a better job in visualizing the model's effectiveness. From the Confusion Matrices, the Precision, Recall, Accuracy and F1 Score can be easily measured. The results of Confusion matrices that are calculated for all the models show that DenseNet121 model has indeed performed really well on the datasets. It also becomes very clear to visualize how adding Batch Normalisation and ReLU activation layers can help to increase model performance.

VII. CONCLUSION

With the experiments conducted for this study on ear datasets, it is finally established that ears are indeed unique for being able to be classified correctly by artificial neural models and also concludes that both ears of individuals can be used to ascertain an individual's identity with no issues whatsoever. This paper has studied image classification upon ears of different individuals and provided with acceptable proof to establish the same. Models of DenseNet121, VGG-19, Inception-V3 and VGG-16 have been exploited for the study and based on the study, DenseNet outperformed the rest with a precision of 95.47%. In future works, these models will be used to develop practical applications for Ear Biometry and further research on upcoming state-of-the-art models upon this field will be conducted to compare computational efficiency along with accuracy in recognising individuals.

VIII. DISCUSSION

In future works, these models will be used to develop practical applications for Ear Biometry and further research on upcoming state-of-the-art models upon this field will be conducted to compare computational efficiency along with accuracy in recognizing individuals.

ACKNOWLEDGEMENTS

The author extends appreciation to all the individuals who provided their hand in collecting the ear images and providing resources for the successful completion of this study.

REFERENCES

[1] A. Iannarelli. Ear Identification. Forensic Identification Series. Paramont Publishing Company, Fremont, Califomia, 1989.

[2] E. Gonzalez, L. Alvarez and L. Mazorra, "Normalization and feature extraction on ear images," 2012 IEEE International Carnahan Conference on Security Technology (ICCST), 2012, pp. 97-104, doi: 10.1109/CCST.2012.6393543.

[3] Ahila Priyadharshini, R., Arivazhagan, S. & Arun, M. A deep learning approach for person identification using ear biometrics. *Appl Intell* 51, 2161–2172 (2021). https://doi.org/10.1007/s10489-020-01995-8

[4] Nejati, H., Zhang, L., Sim, T., Martinez-Marroquin, E., & Dong, G. (2012, November). Wonder ears: Identification of identical twins from ear images. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (pp. 1201-1204). IEEE.

[5] Claes, P., Reijniers, J., Shriver, M. D., Snyders, J., Suetens, P., Nielandt, J., De Tré, G., & Vandermeulen, D. (2015). An investigation of matching symmetry in the human pinnae with possible implications for 3D ear recognition and sound localization. *Journal of anatomy*, *226*(1), 60–72. https://doi.org/10.1111/joa.12252

[6] Yaman, D., Eyiokur, F.I., & Ekenel, H.K. (2020). Ear2Face: Deep Biometric Modality Mapping. *ArXiv*, *abs/2006.01943*.

[7] M. Burge and W. Burger, "Ear biometrics in computer vision," Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, 2000, pp. 822-826 vol.2, doi: 10.1109/ICPR.2000.906202.

[8] Hoogstrate, A. J., Van Den Heuvel, H., & Huyben, E. (2001). Ear identification based on surveillance camera images. Science & justice : journal of the Forensic Science Society, 41(3), 167–172. https://doi.org/10.1016/S1355-0306(01)71885-0.

[9] M. Choraś, "Ear Biometrics Based on Geometrical Method of Feature Extraction," Lecture Notes in Computer Science, pp. 51–61, 2004.

[10] J. D. Bustard and M. S. Nixon, "Toward Unconstrained Ear Recognition From Two-Dimensional Images," in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 40, no. 3, pp. 486-494, May 2010, doi: 10.1109/TSMCA.2010.2041652.

[11] Pflug, Anika and C. Busch. "Ear biometrics: a survey of detection, feature extraction and recognition methods." *IET Biom.* 1 (2012): 114-129.

[12] K. Annapurani, M.A.K. Sadiq, and C. Malathy. 2015. Fusion of shape of the ear and tragus - A unique feature extraction method for ear authentication system. <i>Expert Syst. Appl.</i> 42, 1 (January 2015), 649–656. DOI:https://doi.org/10.1016/j.eswa.2014.08.009

[13] E. H. Said, A. Abaza and H. Ammar, "Ear segmentation in color facial images using mathematical morphology," 2008 Biometrics Symposium, 2008, pp. 29-34, doi: 10.1109/BSYM.2008.4655519.

[14] Li Yuan, Zhichun Mu, "Ear Recognition Based on Gabor Features and KFDA", *The Scientific World Journal*, vol. 2014, Article ID 702076, 12 pages, 2014. https://doi.org/10.1155/2014/702076

[15] Cintas, C., Quinto-Sánchez, M., Acuña, V., Paschetta, C., Azevedo, S., Cerqueira, C.C., Ramallo, V., Gallo, C., Poletti, G., Bortolini, M., Canizales-Quinteros, S., Rothhammer, F., Bedoya, G., Ruiz-Linares, A., González-José, R., &

Delrieux, C. (2017). Automatic ear detection and feature extraction using Geometric Morphometrics and convolutional neural networks. *IET Biom., 6*, 211-223.

[16] Eyiokur, F.I., Yaman, D., & Ekenel, H.K. (2018). Domain Adaptation for Ear Recognition Using Deep Convolutional N eural Networks. *ArXiv, abs/1803.07801*.

[17] Khaldi, Y., Benzaoui, A. A new framework for grayscale ear images recognition using generative adversarial networks under unconstrained conditions. *Evolving Systems* (2020). https://doi.org/10.1007/s12530-020-09346-1

[18] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Pr oceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[19] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[20] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv pre print arXiv:1409.1556*.

[21] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for comput er vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[22] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[23] Zhu, Q.; Mu, Z. PointNet++ and Three Layers of Features Fusion for Occlusion Three-Dimensional Ear Recognition Ba sed on One Sample per Person. *Symmetry* **2020**, *12*, 78.

[24] Ramos-Cooper, S.; Gomez-Nieto, E.; Camara-Chavez, G. VGGFace-Ear: An Extended Dataset for Unconstrained Ear R ecognition. *Sensors* **2022**, *22*, 1752.

[25] Lei, Y.; Qian, J.; Pan, D.; Xu, T. Research on Small Sample Dynamic Human Ear Recognition Based on Deep Learning . Sensors 2022, 22, 1718.

[26] Sandhu, Ramandeep, et al. "Enhancement in performance of cloud computing task scheduling using optimization strate gies." Cluster Computing (2024): 1-24.

[27] kumar Mall, Pawan, et al. "Self-Attentive CNN+ BERT: An Approach for Analysis of Sentiment on Movie Reviews Us ing Word Embedding." International Journal of Intelligent Systems and Applications in Engineering 12.12s (2024): 612 -623.

[28] Narayan, Vipul, et al. "7 Extracting business methodology: using artificial intelligence-based method." Semantic Intelli gent Computing and Applications 16 (2023): 123.

[29] Narayan, Vipul, A. K. Daniel, and Pooja Chaturvedi. "E-FEERP: enhanced fuzzy based energy efficient routing protoco l for wireless sensor network." Wireless Personal Communications (2023).

[30] Mall, Pawan Kumar, et al. "Rank Based Two Stage Semi-Supervised Deep Learning Model for X-Ray Images Classific ation." (2023).

[31] Kumar, Abhay, et al. "Right to education act: universalisation or entrenched exclusion?." Journal of Social Inclusion Stu dies 5.1 (2019): 89-111.