How to Cite:
Sathisha, N., Shivaprasad, K. M., Madhu, C. G., & Vidya, J. (2022). Human ear identification system using shape and structural features based on SIFT and ANN classifier. *International Journal of Health Sciences*, *6*(S1), 5950–5962. https://doi.org/10.53730/ijhs.v6nS1.6216

Human ear identification system using shape and structural features based on SIFT and ANN classifier

Dr. N Sathisha
Assistant Professor, Department of Electronics and Communication Engineering, Govt. S K S J Technological Institute, Bangalore, Karnataka 560001, India,
Email: nsathisha@gmail.com

Dr. Shivaprasad K.M
Professor, Department of Electronics and Communication Engineering, R L Jalappa Institute of Technology, Doddaballapur, Karnataka 561203, India,
Email: shivaprasadkm@rljit.in

Dr. Madhu Chandra G
Associate Professor, Department of Electronics and Communication Engineering, R L Jalappa Institute of Technology, Doddaballapur, Karnataka 561203, India,
Email: madhuchandrag@rljit.in

Mrs. Vidya J
Assistant Professor, Department of Computer Science and Engg, GITAM School of Technology, Bengaluru, Karnataka 561203,India,
Email: vjyothip@gitam.edu

***Abstract*---**People have become more interested in biometrics recognition as technology has advanced. Biometrics is the study of automatic systems for distinguishing people related to physical or behavioural features. Methods for finding favourable biometric traits have acquired more attention in recent years. The ear is among the most consistent biometric traits because it does not alter with aging or emotion. The human ear is an excellent source of data for passive personal authentication. The ear looks to be an excellent potential solution because it is accessible, images are easily obtained, and the ear structure doesn't change over the years. The ear meets a biometric criterion (universality, distinctiveness, durability, and collectability). The biometric field is working hard to develop automated individual identification using ear images. Several face recognition approaches fail in this present COVID-19 outbreak due to the mask-wearing situation. The human ear is attractive data for passive personal

Manuscript submitted: 09 Feb 2022, Manuscript revised: 27 March 2022, Accepted for publication: 18 April 2022

authentication because it does not need the individual's involvement to recognize the ear. In this research, we suggested an automatic ear identification system using Scale Invariant Feature Transform (SIFT) and an Artificial Neural Network (ANN). First, a pre-processing phase in which all images are converted to grayscale and the edge of the ear is detected using the canny algorithm. Following that, SIFT is employed to extract the essential features from the pre-processed ear sample. Finally, using ANN, the obtained characteristics were classified. The experimental findings demonstrated that the above-mentioned approach generates better results, with an overall accuracy of 94% in the training phase and 90% in the testing phase.

***Keywords*---**Ear, Gray conversion, Edge, Feature, Neural Network, Accuracy, Loss.

Introduction

PINs, Log-ins & Passcodes, Identity Cards, and Distinctive Keys are all classic sources of personal authentication that rely on the user's knowledge. The downsides of these strategies include their difficulty to memorize, ease of loss, lack of safety, the ease with which cards and keys are stolen, and the ease with which passwords could be hacked. Due to the shortcomings of conventional methods of identification, it is more practicable to employ biometrics to identify individuals. It is possible to recognize persons automatically depending on their physiological features [1] with the use of biometrics recognition. An example of a physiological biometric is hand geometry, fingerprint, handprints, ear, eye, eyeball, and face identification. Stride, keystroke, and hand vein are all instances of behavioural biometrics, which are biometrics that is based on the behavioural elements of an individual. The ear fulfilled all of the biometric standards since it possesses desirable qualities such as those mentioned above [2]:

- Its universality indicates that everybody should be able to access it;
- Individuality requires that each person have their own.
- The term "permanence" refers to the fact that this quality does not change through time.

Children between the ages of four months and eight years, as well as individuals beyond the age of 70, can demonstrate changes in their ears. The use of the ear as a source of data for human identification provides several advantages, including the following:

- The ear's dependability and robustness can be retrieved from a distance.
- The ear does not change considerably during human life.
- Ear prints could be used to identify a suspect at a crime scene.

To be sure, there are certain hurdles to using one's ear to identify someone, such as having long hair on one's ear, wearing a headscarf to cover one's hair, and the quantity of light available to the person who is wearing it. To verify the person’s identity, an ear biometric system leverages the characteristics of the ear shown in figure 1 to perform identification or verification. The variations in the complicated structure of the ear serve as the basis for this comparative study. During the first four months of life, the ear's growth is influenced by gravity, which can result in the eardrum being stretched out of place. It is feasible to extract reliable and robust features from the ears, even though individuals are unable to recognize one another exclusively based on their ears. Other biometrics, such as fingerprints, hand geometry, and retinal scans, may be perceived as intrusive by users if they necessitate close contact. Iannarelli [3] conducted a study in which he established the viability of ear biometrics. Ear biometrics could be used in a variety of applications such as ATMs, evidence surveillance and identification systems, and restricted admission to any restricted area.

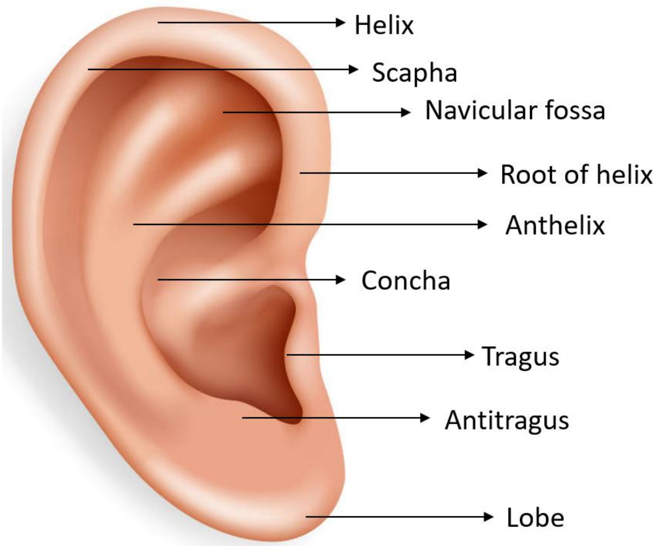

Fig. 1. Structure of ear

Literature survey

To identify a person, the method presented in the paper [4] uses 2D ear imaging. Using the YOLO machine learning (ML) technique, ear images were categorized and their originating individual was identified. The author produced an ear dataset from 164 individuals, totalling 27,592 training images. For testing, 5 images from each of the 164 individuals in the randomized 820-image sample will be used. The training and testing accuracy is 82.5%, 75%. This concept is based on the Python programming language and employs a GPU-based deployment. The research [5] provides a novel architecture for ear recognition based on the learning distance metric (LDM) through Support Vector Machine (SVM) since the LDM and the directed acyclic graph SVM (DAGSVM) have higher generalization abilities than previous ear recognition classifiers. This paper investigates metric learning in the setting of metric SVMs, proposing a hybrid LDM-DAGSVM model for an SVM ear identification system. As an alternative to conventional ear biometric techniques, the proposed solution uses SVM to create a Mahalanobis distance metric to increase variance across classes while decreasing variance within them. By doing trials on complicated ear datasets, it is feasible to achieve better results than current ear identification systems. The suggested technique achieves classification accuracy of 98.79%, 98.70%, and 84.30% for AWE, AME, and WPUT datasets. The classification of ear images will be accomplished by the use of a hybrid approach, as reported in the publication [6]. It makes use of both an RBF network and an Independent Component Analysis (ICA). Instead of employing a feature vector built from grayscale image pixel values, the associated coefficients of these combinations are fed into an RBF network. The RBF adaptability and the ICA's ability to retrieve local information are efficiently combined. As a result, the system's durability has been enhanced. The ICA RBF approach's recognition rate was greatly improved in the experiments.

In the research, deep Residual Networks (ResNet) of varying depths are used to create ear recognition models [7]. Describe three distinct transfer learning strategies for overcoming the difficulties of ear recognition caused by the limited amount of ear images. ResNet topologies or end-to-end system designs are employed as feature extractors. Set the network weights before training the fully connected layer to recognize sounds using pre-trained models trained on a range of visual recognition tasks. The second stage entails optimizing all pre-trained models using the training datasets. This final layer of fine-tuned ResNet models generates feature extractors, which are then classified using SVM classifiers. Finally, to improve the overall performance of the system, create ensembles of networks with varying depths. Extensive tests are performed on the developed models using ear images from the benchmark datasets acquired under controlled and uncontrolled imaging scenarios. As an alternative, the author of the paper recommends Deep Unsupervised Active Learning as an initial training strategy [8]. A classification strategy can gradually obtain various knowledge during the testing stage using the recommended teaching model, obviating the necessity for human direction or judgment correction. During the testing stage, two types of training are used: supervised and unsupervised learning. The model's existing knowledge is continuously expanded during the testing stage, with high-confidence labeling expected. The model's early detection performance must be strong to enhance the proposed method. The author achieved this by using a pre-trained model on three different datasets. By colourisation the images of the USTB2 dataset with a Generative Adversarial Network, this strategy dramatically improved the detection performance of the data. A CMBER-SVM technique based on Chainlet Multi-Band Ear Recognition is described in the research [9]. The suggested method separates the grey image into various bands depending on the severity of its pixels, comparable to a multispectral image. The normalizing channels are next submitted to Canny algorithm, which determines the edges that correlate to each group's ear patterns. To make a single digital edge map, the resulting binary edge mappings are flattened. A histogram is generated for each succeeding four-cell group, and the resultant histograms are standardized and combined to construct a chainlet from the actual image. After that, the dataset's chainlet histogram vectors are utilised to train and validate a coupled SVM. The proposed CMBER-SVM technique exceeds state-of-the-art statistical and learning-based ear recognition systems in tests employing images from two benchmark ear image datasets.

Methods

Ear recognition is very important, but it does not give a high level of accuracy when dealing with image processing. Many challenges are present in image processing techniques. to solve all the issues and increase the recognition rate ML and NN concepts are applied. The overall process of the automatic recognition of the ear is detailed below. The data is very important for research. The benchmark dataset of the ear is collected online. The collected sample is passed through the pre-processing technique. The processed image gives a good result when compared to the raw image. After pre-processing, the important shape and structural features must be extracted from the tan image sample. The SIFT algorithm is chosen for doing the task. Next to feature extraction, the classification algorithm is applied. The ANN is taken for ear recognition. The process flow of ear recognition using SIFT and ANN is shown in figure 2.

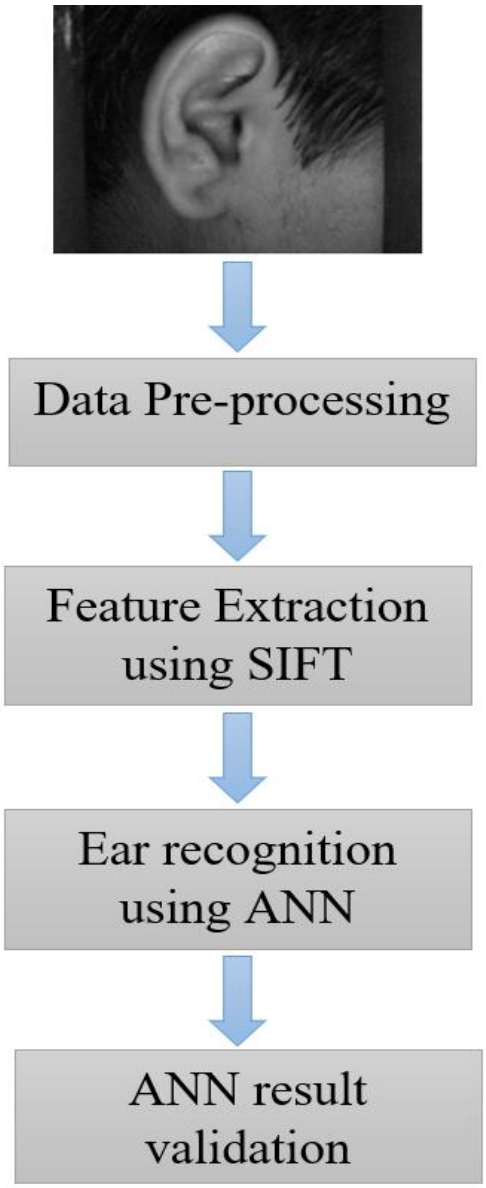

Fig. 2. Ear recognition workflow using SIFT and ANN

Data

Personal identification based on ear images is gaining popularity in the biometrics literature. The lack of a uniform large-scale publicly available ear database complicates the objective evaluation of ear identification technologies. As a result, the Biometrics Research Laboratory at IIT Delhi has been collecting ear image databases from volunteers since October 2006. The IIT Delhi ear image database [10] is made up of ear image databases collected from IIT Delhi students and personnel in New Delhi, India. This database was compiled on the IIT Delhi campus between 10- 2006 and 06- 2007 using a quite easier imaging arrangement. All images were obtained from a distant (touchless) that used a straightforward photographic setup, and the entire process is finished inside the lab. The database now contains 121 persons, each of whom has at least three ear images. All people between the ages of 14 and 58 are included in the database. Each user in the database of 471 images has been allocated a unique numerical identification/number. These images have a resolution of 272'204 pixels and exist as a jpeg type. This database provides automatically normalized and resized ear images of size 50'180 pixels. An enhanced version of the ear database with 754 images of ears from 212 different individuals has been added and is accessible upon request. The sample ear dataset is shown in figure 3.

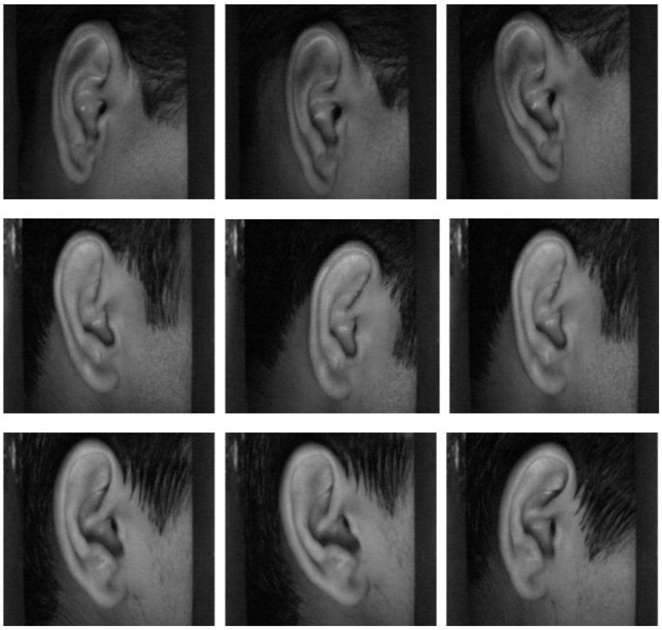

Fig. 3. IIT ear dataset

Data Preprocessing

The image pre-processing is detailed in this section. The pre-processing consists of two major steps namely gray scale conversion and edge detection. The techniques are detailed below.

Gray conversion

Absolute red's RGB encoding is (255, 0, 0), perfect green's is (0, 255, 0), and perfect blue's is (0, 0, 255). The very first integer indicates the value of red, the second digit reflects the quantity of green, and the third digit determines the value of blue throughout all RGB representations. The span of the 3 digits is 0 to 255. Gray-scale images are available in white, black, and a variety of grey tones. Every grey value is a single integer between 0 to 255. Converting the full-colour image to an 8-bit representation is the next step. This minimizes the complexity of the space, allowing for faster additional research without jeopardising reliability. At times, three-dimensional spatial coordinates are used. Grayscale images are images that contain no information other than their intensity [11].

Edge detection

Many people believe that the Canny technique is good for edge detection [12]. The intelligent edge detection algorithm extracts the inner and outer layer edges of the ear from the acquired image. This improves the sharpness of the ear. To identify and validate the image, utilise this method to determine the exact bounds of the ear. The Canny Edge Detection technique operates in five steps. It is described in greater detail below.

1. Smoothing: When an image is captured using a camera for the first time, it contains some noise data. In the early phases, a Gaussian filter is utilised to smooth the image.
2. Gradient identification: The gradients in the image can be utilised to pinpoint these issue regions. After the smoothing procedure, use the Sobel operator to find each pixel's gradient points. The first stage is to evaluate the gradient in both the x and y directions using kernels. The Pythagorean Theorem can be used to compute edge strengths, which are also another term for gradient magnitudes.
3. Non-maximum suppression: Sharpening can be used to improve the appearance of images with blurry gradients. All local maxima must be retained, and the image must be cleaned up in the following steps: To smooth out the gradient to go to the nearest coordinate, utilise an eight-connected neighbourhood. Comparing the pixel edge strength in the optimistic and pessimistic gradient directions is conceivable. Maintain the current pixel's edge strength value if it is high; otherwise, lower it.
4. Double Thresholding: The pixel-by-pixel strength of the edge pixels that remain after the non-maximum suppression phase is astounding. Even if the rough surfaces on some of those pixels are actual edges, they could introduce noise or colour changes into the image. Threshold settings can be used to separate these pixels, retaining just the sharpest edges. When the threshold is set to a high value, these edge pixels are strong; when the threshold is set to a low value, they are weak.
5. Edge tracking by hysteresis: BLOB-analysis can be employed to track an edge. It uses an 8-connected neighbourhood to link to BLOBs. BLOBs with minimal sturdy edge pixels are maintained, whereas other BLOBs are hidden.

SIFT Algorithm

The SIFT [13] has addressed a wide range of practical difficulties through the use of low-level extracting features and image matching. A limitation of the early Harris operator is that it is extremely sensitive to fluctuations in image scale, making it inadequate for matching images of varying sizes. For the SIFT transform to work, two stages must be completed: extraction of features and feature characterization. The description step will be focused on the low-level characteristics that are employed in object matching; a more in-depth analysis will take place later. The low-level feature selection technique used by the SIFT method retrieves the most significant properties from an image in a way that is irrespective of the image scale rotation and brightness levels. Additionally, the formulation reduces the likelihood of suboptimal retrieval as a result of occlusion congestion and deformation during the retrieval process. This also highlights how many of the previously described strategies can be merged and used to their full capacity. First, the difference of the Gaussian operator’s operator is used in an image to discover potentially relevant properties. The approach attempts to ensure that feature selection is independent of feature size or direction. The local gradient direction is being used to evaluate the features to determine the orientation. After that, a feature representation is created that is equipped to deal with changes in light as well as local form deformities in the real world. The operator's primary contribution is the application of local experience to the data generated by conventional operators. Because they are outside the scope and goal of this paper, it is preferable to leave the operations to the source. Figure 4 depicts the features detected by SIFT. The amplitude and orientation of the detected relevant traits are shown by the width and orientation of the white lines, respectively. The major elements are completed with the lizard's head and a few leaves. The smaller white lines in this figure show minor traits that are grouped around background elements. There are a lot littler details in this image's complete collection of features, notably in the image's textured areas. We'll go through how this can be used in shape extraction later, but for now, let's focus on the fundamental low-level features.

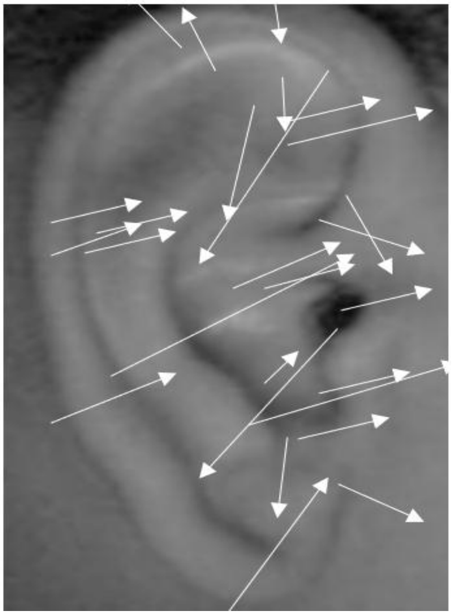

Fig. 4. Detecting features using SIFT

ANN Model

ANN is a subset of ML that has grown in importance in recent years in terms of research and development. Through the concept of ML, a computer can learn to understand the structure of its incoming data by using a mathematical or statistical model. ANN, on the other hand, is not a straightforward technique; they require extra steps and must deal with patterns that are tightly connected with massive amounts of data. This requires a deep understanding of the technique as well as the development of a systematic algorithm. ANN is inspired by the biological human brain, which is made up of up to 60 trillion interconnected neurons that execute network patterns of decision making. Based on this core concept, ANN begins with simple, interconnected neurons that act as a single processor. Based on McCulloch and Pits' neuron model [14], the notion of perceptron was formed. An ANN's basis is made up of a single layer of input, process, and output parts. From the most fundamental information processing cycle notion, a sophisticated mathematical formulation can be developed to offer the best potential solution for any dataset or issue segment. To complete a network cycle, neurons must be tested using feed-forward and backward algorithms. In the past and currently, many scholars have addressed and investigated backward algorithms and backpropagation. The architecture of ANN is shown in figure 5.

The number of layers at the beginning is usually determined by the difficulty of the problem to be solved. Iterating through four unique NN phases is required for the classification job: initiation, activation, weight tuning, and repetition. The phases of activation functions, however, fluctuate depending on the problem to be solved. Neurons are the basic information processing pieces of a NN, and they are linked together by connections. Each link had its weight, and each neuron had numerous weights in addition to its new, altered weight. From the input neuron, links will expand outward. This sort of NN is known as a feed-forward network. Each link has a number and weight associated with it that connects it to the output. The term "perceptron" refers to a network of nodes that are linked together. After the input connections have been individually weighted, they will be merged to form an activation function. CG formulae were used in this step to determine the optimal weight related to the next layer; layers 1 to 2, 3, or up to layer n. The number of layers in a particular network is determined by the complexity of the problem to be solved. The following NN process diagram depicts the feedforward and backpropagation processes.

As indicated in the following graphic, an ANN technique was outlined by adding a CG interpretation for every input and proceeding with the same composition into the next layer of the ANN. Because the purpose of the procedure is to select the ideal output value depending on a specific input value, error correction will be disregarded at this level.

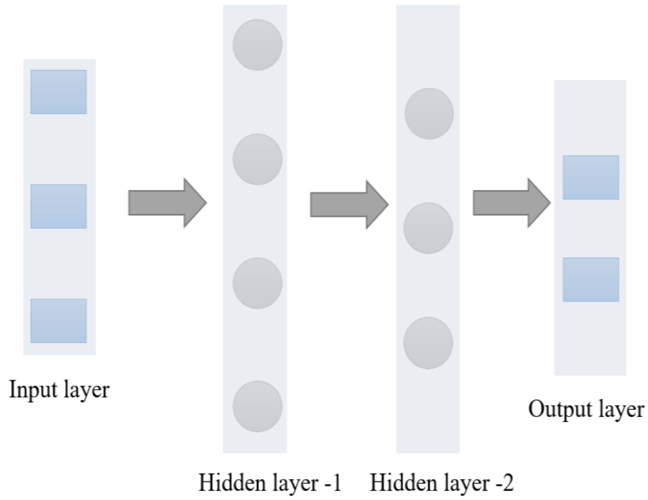

Fig. 5. ANN Architecture

Result and Discussion

The result obtained from ANN is discussed in this section. The collected image from the IIT database is processed for converting RGB to gray image and the edge of the ear is detected using the canny algorithm. Before giving the image directly to the ANN the features of the ear are extracted from the processed image using SIFT. The sample collected is separated into two modes. In training mode, 80% of the image sample is used and in test mode remaining 20% of the image, the sample is used. The training and test data are used parallelly to train and validate the ANN model. The ANN model quality is verified using the metrics terms such as accuracy and loss. The accuracy is a positive metric and loss will be a negative metric. For a good model, the positive metrics should be high and the negative should be below. Figure 5 shows the accuracy obtained by ANN in training and validation mode. The blue plot illustrates the accuracy obtained by ANN on training data in each epoch. Similarly, the orange plot shows the validation accuracy of ANN from 0^{th} epoch to 15^{th} epoch. After 10^{th} epoch, the accuracy level will become stable and nearer to 90%. Finally, at the last epoch, the accuracy obtained by training mode is 94% and by validation is 90%.

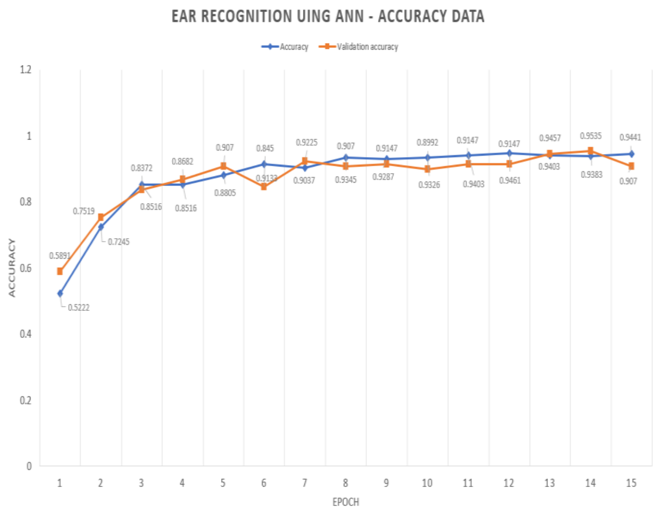

Fig. 5. Training accuracy and validation accuracy of ANN

Figure 6 depicts the loss achieved by ANN during training and validation. The blue plot displays the ANN training loss in each epoch. Similarly, the orange plot depicts the validation loss of ANN from the first to the fifteenth epoch. After the tenth epoch, the lower response will stabilize and approach 30%. Finally, at the last epoch, the training mode's loss was 19.57%, while the validation mode's loss was 27.18%.

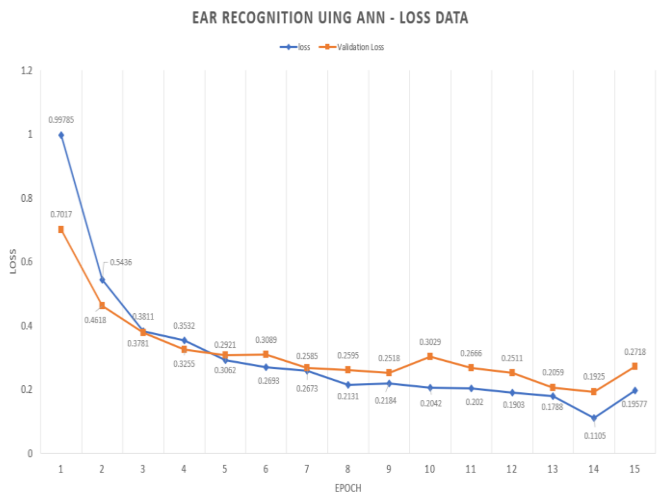

Fig. 6. Training loss and validation loss of ANN

Conclusion

Researchers in computer vision and intelligent systems have been working tirelessly to create a dependable automated system for validating one's identity.

Biometrics are commonly utilised in human identification systems due to their long-term stability, ease of acquisition, and individuality. In recent years, there has been a lot of interest in detecting people by their ears. In addition to facial recognition technologies, which have struggled in real-world settings, it represents a significant advancement in biometric research. This is because there is a lot of variation in shape and lighting conditions, as well as a planar depiction of an item that changes in profile. To identify a certain individual, ear identification relies on the effective form of the ear's inner and outer layers. Texture and ear heredity are lost as a result of traditional methods. Edge detection, SIFT feature extraction, and ANN can all be used to improve ear recognition accuracy. Using the intended technique in conjunction with face-recognition technologies will improve future security. Because it is impossible to validate an image of someone who only shows one side of his or her face. The authentication process can be accomplished with the help of the ear and a section of the face. It is less priced and outperforms other solutions.

References

[1] A. S. Anwar, K. K. A. Ghany and H. ElMahdy, "Human ear recognition using SIFT features," *2015 Third World Conference on Complex Systems (WCCS),* pp. 1-6, 2015

[2] M. Burge and W. Burger, "Ear Biometrics", A.K. Jain, R. Bolle, S.Pankanti (Eds.): in Biometrics: Personal Identification in Networked Society, *Springer*, pp. 273-286, 2013.

[3] A. Iannarelli, "Ear Identification", *Forensic Identification Series, Paramont Publishing Company, Fremont,* 1989.

[4] S. Hossain and S. Akhter, "Realtime Person Identification using Ear Biometrics," *2021 International Conference on Information Technology (ICIT)*, pp. 104-108, 2021

[5] I. Omara, G. Ma and E. Song, "LDM-DAGSVM: Learning Distance Metric via DAG Support Vector Machine for Ear Recognition Problem," *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1-9, 2020

[6] Hai-Jun Zhang, Zhi-Chun Mu, Wei Qu, Lei-Ming Liu and Cheng-Yang Zhang, "A novel approach for ear recognition based on ICA and RBF network," *2005 International Conference on Machine Learning and Cybernetics*, vol. 7, pp. 4511-4515, 2005

[7] H. Alshazly, C. Linse, E. Barth, S. A. Idris and T. Martinetz, "Towards Explainable Ear Recognition Systems Using Deep Residual Networks," in *IEEE Access*, vol. 9, pp. 122254-122273, 2021

[8] Y. Khaldi, A. Benzaoui, A. Ouahabi, S. Jacques and A. Taleb-Ahmed, "Ear Recognition Based on Deep Unsupervised Active Learning," in *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20704-20713, 2021

[9] M. M. Zarachoff, A. Sheikh-Akbari and D. Monekosso, "Ear Recognition using Chainlet based Multi-Band SVM," *2021 IEEE International Conference on Imaging Systems and Techniques (IST),* pp. 1-5, 2021.

[10] https://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Ear.htm

[11] Kumar, Tarun & Verma, Karun, "A Theory Based on Conversion of RGB image to Gray image", *International Journal of Computer Applications,* vol. 7, 2010

[12] Ramnarayan, Nikita Saklani, Vasundhara Verma, “A Review on Edge detection Technique-Canny Edge Detection”, *International Journal of Computer Applications,* vol. 178, no. 10, pp. 0975 – 8887, 2019

[13] Lowe, D. G., “Object Recognition from Local Scale-Invariant Features”, *Proceedings of ICCV*, pp 1150-1157, 1999.

[14] Negnevitsky M. “Artificial Intelligence: A Guide to Intelligent Systems”. *Addison Wesley: Longman Publishing*, 2009