International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Transfer Learning with Optimization Enabled Person Recognition using Ear Images

Sowmya M N*[1], Dr. Keshava Prasanna[2]

Submitted: 06/12/2023 **Revised**: 12/01/2024 **Accepted**: 30/01/2024

Abstract: The human ear is a good source of data for passive human recognition as it doesn't need the assistance of the individual whose identity we're trying to determine and because the ear's structure doesn't vary significantly over time. A hybrid method for person detection utilizing ear pictures called Gannet Sparrow Search Optimization _enabled Convolutional Neural Network with Transfer Learning (GSSO_CNN-TL) is created to counteract these constraints. The input ear image is initially attained from an ear recognition dataset and fed to pre-processing stage, where external noise are removed using Non-Local Means (NLM) filter. After that, techniques including rotation, flipping, cropping, and colour augmentation are applied to the pre-processed image. Then features like Scale-invariant feature transform (SIFT), Speeded up robust features (SURF), Pyramid Histogram of Oriented Gradients (PHoG), and Local Binary Pattern (LBP) are retrieved. Lastly, person identification is done using CNN with TL, where CNN is employed with hyper parameters from the trained models, namely deep batch-normalized eLU AlexNet (DbneAlexNet). Nonetheless, the GSSO is efficiently used to train classifier. By merging Gannet Optimization Algorithm (GOA) and Sparrow Search Algorithm (SSA), the created GSSO is produced. Furthermore, the devised approach achieved maximum precision, recall, and f-measure of 92.6%, 95.8%, and 94.2% respectively.

Keywords: *Convolutional Neural Networks, Ear image, Person recognition, Transfer learning*

1. Introduction

The human ear is regarded as significant biometric as it exhibits many distinctive characteristics. Researchers discovered that even identical twins have varied ear shapes [9]. The simplicity of ear image analysis, invariance to facial expressions, and stability over time make them superior to other biometric features including faces, iris, fingerprints, and palm prints [10]. Owing to these advantages, ear pictures can be utilized for many different things, including biometric identification, clinical asymmetry research, genetic link analysis, and gender recognition [11] [3]. Ear identification is a technique for automatically recognizing human based on the physical characteristics of their ears. It is applicable to a wide range of application domains, such as surveillance, unlocking user's devices, forensics, and identity verification, and.Based on the unique structure of the ear shape, an effective recognition system can be created employing ear photos as a rich source of biometric data. Additionally, the human ear has a number of advantageous qualities, such as its ability to be easily captured from a distance, its stability over time, its capacity to distinguish between identical twins [9], and its insensitivity to emotions and facial expressions [12], [13] [4]. Due to its non-intrusive nature and low user involvement during recognition operations, the ear biometric is a good replacement for other widely used physiological modalities including fingerprint, iris, and palm print. To record 2-Dimensional ear images using this biometric, which has a promising performance over face recognition, a low-cost camera is needed. Once more, due to the smaller size of the ear shape, it can be combined with another biometric trait with ease to address the issues with unimodal biometric systems [5].

As a result of its consistency across time, ease of acquisition, and individuality, biometrics is used by the majority of human identification systems. Several biometric research projects have been put into practice effectively and are now open to the public. The biometric methods are most frequently used for person identification. The majority of biometrics used for human identification depends on the cooperation of the matching individual to obtain the biometric features [1]. Biometrics is a crucial part of cyber security that allows for the identification of people by confirming their physical and behavioural characteristics. For authenticity verification, it is currently the most precise and efficient physical security mechanism in use. Based on their inherent behavioural or physical characteristics, each person can be correctly identified via biometric-based authentication. Researchers [14] have described biometric systems that use physiological characteristics including the face [15], fingerprint, iris, palm prints, and ear. The biometric security system scans a person who attempts to enter, evaluates their attributes, and compares them to previously stored data. When a match is made, access to the space or tools is granted to the person. Touch-based and touchless biometrics falls within this category. One of the most common touch-based biometrics now in use is the fingerprint. Hackers can now access crucial information by imitating fingerprints, due to technological advances [3].

In several application fields, including image classification, object identification, and biometric recognition [16], deep learning techniques, more notably Deep Convolutional Neural Network (DCNN), have recently made significant strides. These advancements are outcome of a number of factors, including the accessibility of enormous amounts of labelled data, potent hardware (such as Graphics Processing Unit (GPUs)) for accelerating computations, well-designed deep network architectures, efficient optimization methods, and technical

[1] *Research Scholar, Channabasaveshwara Institute of Technology, Gubbi, Visvesvaraya Technological University, Belagavi*
ORCID ID : 0009-0008-6938-3018
[2] *Research Supervisor,Channabasaveshwara Institute of Technology, Gubbi, Visvesvaraya Technological University, Belagavi*
ORCID ID : 0000-0001-7166-3555
** Corresponding Author Email: sowmyamn@sit.ac.in*

advancement in training deep networks. By thoroughly training the system from start to finish, DCNNs are able to extract and classify features without the aid of humans. They are also scalable supervised learning approaches. Nonetheless, optimizing a sizable number of trainable parameters (millions) and sizable labelled datasets are necessary for training DCNNs. Moreover, gathering such large volumes of data can be costly for some real-world applications, which would limit the tremendous potential of deep models [4]. In some applications or areas where data is sparse or expensive to collect, deep models and powerful CNN-based representations are not fully utilized. In forensics and security applications, ear recognition is one of these areas that has potential value. As there aren't enough large-scale labelled ear databases, ear identification technology is limited compared to other biometric-based technologies like face recognition [7]. The deep learning variation makes use of machine learning techniques. CNN and its offspring Neural Network (NNs), Visual Geometry Group Network (VGGNet), Region with CNN (R-CNN), etc. play a crucial role in object identification and classification applications [2].

The major objective is to present an approach, namely GSSO_CNN-TL for precisely identifying a person from ear pictures. The input ear images are primarily taken from ear identification dataset and sent to image pre-processing step. In this case, the pre-processing step involves removing unwanted noise from the input images using the NLM filter. Following that, image augmentation is carried out. The features are retrieved from the augmented image. The person is also recognized using CNN-TL where CNN is combined with hyper-parameters from learned models like DbneAlexnet, which is trained using GSSO.

This work's main contribution is described below.

- Established GSSO_CNN-TL for person recognition using ear images: GSSO_CNN-TL, a powerful technique that makes use of images of ears to identify people. In this case, person identification is done by combining CNN-TL and hyper-parameters from previously trained models like DbneAlexnet, which is trained using GSSO. The combination of GA and SSO is also known as designed GSSO.

2. Motivation

Due to its distinct form and wealth of regional traits, the ear biometric has recently attracted a lot of attention for human recognition. Although it is a difficult process in the context of lighting fluctuations, poor contrast, noise, and variations in position, extracting discriminative features from photos of ears is essential. Some deep learning methods that were previously used took a lot of time and were only effective for small numbers of images. As a result of researching existing techniques for person recognition, a strategy was developed using this approach. Below are outlined the examined research for person recognition, along with limitations.

2.1 Literature survey

AhilaPriyadharshini, R., et al. [1] developed six-layer DCNN architecture for ear recognition. As there are fewer layers in this model, it was not affected by the vanishing gradient problem when bounded activation function tanh was used. Yet, the technique failed to take into account correct illumination and random noise. Hossain, S., et al. [2] designed two techniques namely non-deep Machine Learning (ML) models and Deep Learning (DL)-based ML models to identifying a person from 2D ear images. Here, suggested application for the ML approach was tested. Furthermore, it offered a setting for in-the-moment experimentation. The YOLO performance in real-time classifications was not enhanced by this strategy. Mursalin, M., et al. [3] devised you morph once (YMO) approach for human recognition. This approach lowered computing time in this case by eliminating one-to-one registration between the probe and the gallery, which determined the distance between the parameters stored in the gallery and the probe. However, approach required more effective datasets for better recognition. Alshazly, H., et al. [4] designed an efficient method named transfer learning with DCNN for ear recognition. In this situation, the network with specially scaled inputs was created to preserve the aspect ratio of ear images and produced better results. Nevertheless, this method failed to consider less time during processing.

Sarangi, P.P., et al. [5] suggested a computerized improvement technique called enhanced Jaya Algorithm (EJA) that uses meta-heuristic optimization to improve ear images. In order to enhance the performance of the ear biometric system, EJA implemented a mutation operator in this case with the objectives of speeding convergence and enhancing the quality of ear images. Unfortunately, existing method was unable to be enhanced by including a multi-objective meta-heuristic optimization method. H. Sinha., et al. [6] developed a deep learning strategy for ear localization and recognition. In this case, the framework combined Histogram of Gradients (HOG) with support vector machines (SVM) for ear localization and then used CNN for ear recognition. Unfortunately, this method had the disadvantage of poor lighting. H. Alshazly., et al. [7] created an ensemble of DL models and TL for ear recognition. The produced ensembles in this instance had the highest recognition accuracy across all datasets. Nevertheless, trials were not extended to investigate other DCNN architectures and appropriate learning methodologies. D. Yaman., et al. [8] developed a deep neural network model for person identification accuracy. By using a large-scale dataset with substantial appearance diversity, the accuracy of the reconstructed face images and the model's generalizability were enhanced. Nevertheless, this strategy failed to collect a large-scale dataset of ear-face image pairs in the wild in order to capture additional appearance variants.

2.2 Major challenges

Using ear pictures to identify people has encountered a number of challenges, some of which are listed below:

- ✓ The architecture of a six-layer DCNN for ear recognition was developed in [1].The method's performance was only assessed utilizing restricted databases; conventional evaluation parameters were not used.
- ✓ In [4], a powerful technique for ear recognition called TL with DCNN was developed. This approach overlooked the need for faster processing times.
- ✓ Developed a deep learning method for ear localization and identification in [6]. Unfortunately, in images with different poses, this classifier failed to recognize ears.
- ✓ In [7], an ensemble of DL models and TL was devised for ear recognition. Due to the failure of this strategy to control CNN attention, the networks were unable to

improve their ability to learn by focusing on certain discriminative regions in the input image.

- ✓ When ear images are taken in unconstrained situations, where many possible variations in look and lighting must be taken into account, it can be challenging to accurately identify people.

3. Proposed GSSO_CNN-TL for human recognition using ear images

Initially, the input ear image is acquired from specific dataset represented in [17] and it is fed up into the pre-processing stage, where the external noises in the image is completely discarded using NLM filter [18]. After that, image augmentation is done over pre-processed image using techniques, like rotation, flipping, cropping, and color augmentation. Once the image augmentation is done, features like SIFT [22], SURF [24], PHoG [23] and LBP [25]is extracted. The final step in person identification uses CNN with TL, which combines CNN with hyperparameters learned from trained models like DbneAlexNet. However, the network classifier is effectively trained based on proposed GSSO. The devised GSSO is obtained by the integration of GOA [19] and SSA [20]. Figure 1 displays flow diagram of proposed GSSO_CNN-TL for person identification using ear images.

3.1 Image acquisition

At the first stage in the processing for person recognition, the ear images are retrieved from the ear detection and recognition dataset [17]. Following that, these ear images can be pre-processed. The dataset with ear images is listed below;

$$\varepsilon = \{V_1, V_2, ..., V_f, ..., V_k\} \quad (1)$$

If ε is a dataset containing many ear images, V_f indicates an input ear image that can be pre-processed, and V_k is the image that arrives at position k, where k represents total number of ear images in database.

3.2 Image pre-processing

The pre-processing stage of person recognition is the following step, where V_f is the image fed at this phase. Before enabling the usage of this filtered image for image augmentation, the NLM filter [18] is utilized for pre-processing.

3.2.1 Pre-processing using NLM filter

The nonlocal means filter in [18] lowers pixel noise by replacing a noisy pixel with a weighted average of non-locally nearby pixels throughout the whole image. The NLM filter eliminates noise and sharpens edges without significantly affecting fine structure or details. The weighted average of the image utilizing all of the image's pixels is calculated by applying the equation,

$$\hat{\chi}(X) = \frac{\sum_r \exp\left(-\frac{\| T_X - T_r \|^2}{t^2}\right) \chi(r)}{\sum_r \exp\left(-\frac{\| T_X - T_r \|^2}{t^2}\right)} \quad (2)$$

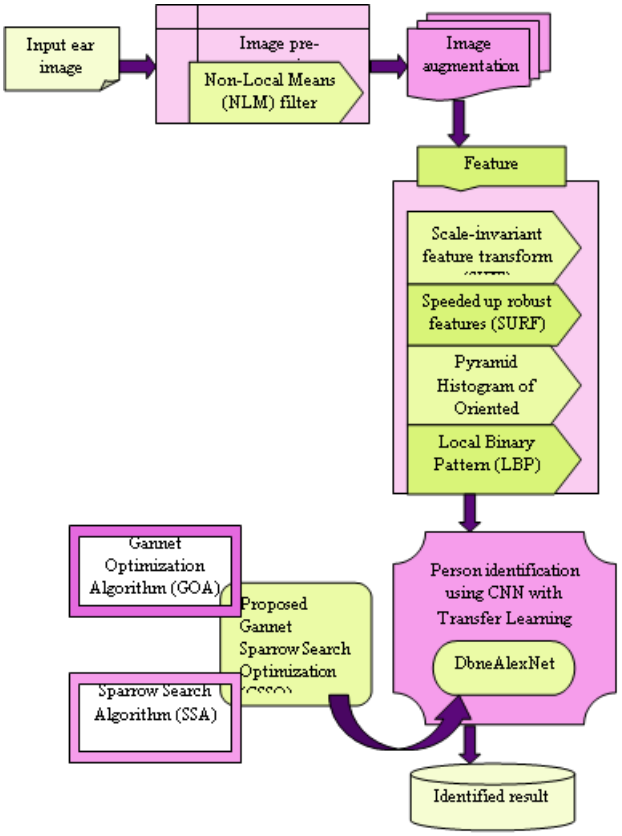

Figure 1. Block diagram of proposed GSSO_CNN-TL for person identification using ear images

Here, the column vectors T_X and T_r, which stand for the pixel colors within blocks centered at X and r, correspondingly. Moreover, gaussian parameter is represented as t. The weighted average gives $\chi(r)$ a bigger weight when block T_r *and* T_X are more similar to one another. A three-dimensional vector in the Red, Blue and Green (RGB) colour space represents the colour $\chi(r)$ of each pixel in this paper. Here, the filtered image output is symbolized as V_λ.

3.3 Image augmentation

This approach artificially increases the size of a database to increase forecast accuracy and allay privacy concerns. Moreover, image augmentation [21] prevents memory utilization and helps to expand the database's capacity.The filteredimage V_λ is given into the image augmentation process, which includes steps like image cropping, flipping, rotation, and hue image.

3.3.1 Image cropping

Cropping [21] is the procedure of selecting at random a portion of an original image. The region is then adjusted to fit the original image. It is a component of the primary image, which is resized as necessary. Random cropping is another name for this process. Cropping represented as ψ_a

3.3.2 Image flipping

Either the horizontal or vertical flip [21] of the image is possible. Images are created by rotating an image a multiple of 90 degrees. Nonetheless, there are several frameworks that do not allow vertical flipping. So must conduct a horizontal flip after turning

the image by 180 degrees to perform a vertical flip. Flipping is specified as ψ_b.

3.3.3 Image rotation

The image can be rotated at 90-degree angles or at smaller angles, depending on the requirements [21]. When the image is rotated at 90-degree angles, no additional background noise is applied after orientation. If rotation is carried out at tiny angles, the opposite is true. Furthermore, if the background is black or white, recently added colour in the form of noise will mix into the image. Here, image rotation is represented as ψ_c.

3.3.4 Hue image or color augmentation

While it is not possible to directly change a pixel's position, it is possible to change its value.It deals with changing the colour properties of an image by changing the values of the pixels [21].Grayscale, contrast, brightness, hue, and saturation adjustments are made to achieve the colour augmentation. Furthermore, hue image is signifiedas ψ_d.

Thus, the process of image augmentation is described as follows:

$$\Upsilon = \{\psi_a, \psi_b, \psi_c, \psi_d\} \tag{3}$$

3.4 Feature extraction

The procedure of extracting features is done by acquiring data-augmented vector Υ. The determination of a number of features, including SIFT [22], PHoG [23], SURF [24], and LBP [25].

3.4.1 SIFT

Object recognition in general led to the development of the SIFT [22]. Its efficacy has recently been demonstrated through its application in biometric recognition systems.Along with affine image alteration, SIFT also transforms image data into scale-invariant coordinates for local properties that are not affected by rotation, scaling, or some variations in lighting. The subsequent procedures are necessary for the development of image feature sets.

i)Scale space extrema detection: Image locations and all scales are identified in this step. To effectively find possible interest sites that are scale-invariant, a Difference-of-Gaussian function is used. A function $L(a,i)$ is used to determine the scale space of an image. It is created by convoluting an input image $g(a,i)$ with a variable scale Gaussian function $D(a,i,\varsigma)$ as specified in equation (4), where ς is the scale of blurring.

$$Y(a,i,\varsigma) = D(a,i,\varsigma) * g(a,i) \tag{4}$$

Equation (5) yields the 2D Gaussian kernel**.**

$$D(a,i,\varsigma) = \frac{1}{2\pi\varsigma^2} m^{\frac{-(a^2+i^2)}{2\varsigma^2}} \tag{5}$$

ii) Key-point localization: To identify location and scale at each prospective place, a thorough model is fitted. Based on indicators of their stability, the important points are chosen.

iii) Orientation assignment: Using the local image gradient directions, a consistent orientation is allocated to each key-point position.

iv) Key-point descriptor: The local image gradients at the selected scale are determined by measuring the region around each key point. They develop into a representation that permits a sizable amount of local shape and illumination modification. The SIFT feature is represented as Q_1

3.4.2 PHoG

The spatial shape descriptor PHOG [23] has recently been applied to the classification of images. The Histogram of Gradients (HOG) is added to each of the smaller segments, preserving the spatial organization of the original image, after the image has been broken into smaller pieces at varying resolutions. A four-level dimensional pyramid is then created after a sophisticated edge detector uses a grayscale image to identify the PHOG feature. However, construction of PHOG, requires that HOG be checked for each bin in each level and added. PHoG feature is signifies as Q_2.

3.4.3 SURF

The SURF [24] algorithm offers a method that can locate important spots inside a certain sort of image, such as corner, edge, and highlight points, and then identify between the various types. Here, Hessian matrix at scale $\wp$ is the image's point M.

$$I(M,\wp) = \begin{bmatrix} P_{nn}(M,\wp) & P_{np}(M,\wp) \\ P_{np}(M,\wp) & P_{pp}(M,\wp) \end{bmatrix} \tag{6}$$

Here, convolution outcome of Gaussian second derivative is denoted as $P_{nn}(M,\wp)$. SURF feature is indicated as Q_3.

3.4.4 LBP

In reality, LBP [25] specifies 3×3 neighbors, offering 8 bit-codes in terms of neighbors across the pixel in the centre. The decimal model for the resulting LBP for the pixel at (w_h, e_h) is:

$$J_1 = LBP(w_h, e_h) = \sum_{\eta=0}^{7} H(j_\mu - j_h)2^\eta \tag{7}$$

Where, in contrast to w_h and e_h, which denote the grayscale values of the centre and surrounding pixels, η denotes the centre pixel's eight neighbors. The LBP feature is represented as Q_4.

$$\phi = \{Q_1, Q_2, Q_3, Q_4\} \tag{8}$$

3.5 Person identification using ear images

Once the feature vector ϕ is obtained, CNN [30] is carried out for the person recognition using ear images. The parameters of the CNN are initialized using the TL concept. Here, the feature vector ϕ is fed into the DbneAlexNet [31] with the purpose of training it to learn the hyper-parameters. These hyper-parameters

are used to change the CNN's settings. The GSSO algorithm, which was created by extending the GOA [19] technique using the SSA [20], is also used to optimize the CNN's weights and bias. The sections that follow provide more information on the recognition procedure, CNN structure, and suggested optimization algorithm.

3.5.1 DbneAlexNet with CNN based TL

The CNN [30] parameters are used for person recognizing, and they are updated using the transfer learning [32] principle, which makes use of the information of hyper-parameters acquired by the DbneAlexNet [31] for training the CNN parameters. The developed feature vector ϕ is used to train DbneAlexNet, and the resulting hyper-parameters of the DbneAlexNet are used to train the CNN. The problems of imbalanced and sparse data are successfully addressed through transfer learning, which also lowers the False Positive Rate (FPR). Figure 2 provides a representation of the transfer learning process.

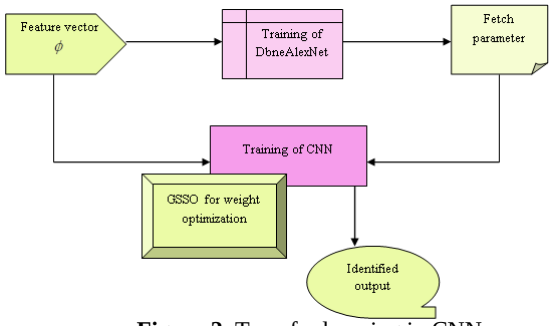

Figure 2. Transfer learning in CNN

a) DbneAlexNet

The DbneAlexNet [31], which is employed for taking images of person recognition, estimates the likelihood of person recognition. There are several different kinds of layers in it, including the batch normalization layer, eLU activation layer, and dropout layer. This $\mathrm{Re}LU$ function allows for faster learning than learning with the function because the activation layer's negative values enable it to step up the activations of average units closer to zero.In neural networks, the internal confounders shift problem develops as a result of changes in the distribution of input data brought on by a change in the number of variables in the previous layer. Batch normalization is a technique for fixing this issue. Another mechanism in this technique is dropout. It is referred to as a method for standardizing CNN models. This method makes it possible to neglect neurons that were randomly selected during training.

b) CNN

There are five layers that make up a CNN: an input layer, a convolutional layer, a pooling layer, an FC layer, and an output layer[29].The convolutional layer's task is to acquire the data and extract its features. It consists of several convolution kernel layers, each of which is associated with a weight and a deviation coefficient. It is possible to express the convolution process as:

$$\gamma_u = l(q_u \otimes \gamma_{u-1} + E_u) \tag{9}$$

Convolutional layer u receives γ_{u-1} as its input, with weight coefficient expected to be q_u and deviation quantity supposed to be E_u. Moreover, the convolution operation is represented as $\otimes$.

Additionally, convolutional layer's activation function is the ReLU function. The sigmoid, tanh, and other activation functions are more difficult to derive than the ReLU activation function. This makes it faster to train models and more effective at preventing gradient emergence. The function of ReLU is expressed as,

$$\mathrm{Re}LU(\gamma_u) = \begin{Bmatrix} \gamma_u & \gamma_u > 0 \\ 0 & \gamma_u \le 0 \end{Bmatrix} \tag{10}$$

By removing redundant information by down sampling, the pooling layer's primary purpose is to achieve invariance and lessen the complexity of the CNN. There are two major approaches to pooling: average pooling and maximum pooling. Max pooling refers to selecting the largest value in the calculation region as the pooling result, whereas average pooling refers to using average value in the calculation area. The below expression indicates max pooling.

$$B_C = Max(R_C^0, R_C^1, R_C^2, R_C^3, ..., R_C^\rho) \tag{11}$$

Here, output of the pooling region C is represented as B_C, and max pooling operation is indicated as Max.

Throughout CNN, FC layers serve as "classifiers." By weighing and balancing those features, they are primarily in responsible for re-mapping the characteristics of the convolutional and pooling layers from the hidden layer space to the sample marker space. In order to avoid over fitting, a corresponding dropout process is set up in the FC layer to randomly remove neurons. The output of CNN architecture is signified as v.The developed GSSO method also modifies the bias and weights of the CNN. Figure 3 displays the structural view of CNN architecture.

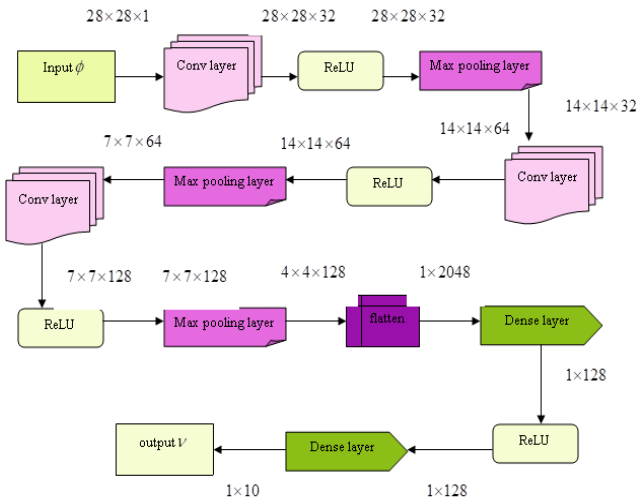

Figure 3. Structural view of CNN architecture

3.5.2 Training procedure for proposed GSSO algorithm

In this paper, person recognition using ear images is accomplished using the GSSO technique, which was created by fusing the GOA [19] and SSA [20] algorithms. Any optimization problem can be solved using the meta-heuristic GOA [19] algorithm, which was inspired by nature. In order to facilitate exploration and exploitation, the GOA mathematically, represents the different distinctive behaviors of gannets when foraging. GOA offered a better solution and ran more quickly in high dimensions. Five engineering optimization issues are addressed with the GOA. The primary benefit of this method is that it has good processing power for problems with high dimensions. Moreover, the SSA [20] takes inspiration from the collective intelligence, foraging, and anti-predator actions of sparrows. In numerous search spaces, SSA performs well. The primary benefit of this SSA strategy is its strong capacity to investigate potential local optimum regions, effectively avoiding the local optimum problem. A combination of GOA and SSA is employed to enhance the entire algorithmic process. The developed GSSO includes the subsequent steps.

Step i): Initialization step

When the GOA starts with a set of random solutions like those in equation (12), the best answer is regarded as the best overall solution.

$$\hbar = \begin{bmatrix} \aleph_{1,1} & \cdots & \aleph_{1,c} & \cdots & \aleph_{1,Dim-1} & \aleph_{1,Dim} \\ \aleph_{2,1} & \cdots & \aleph_{2,c} & \cdots & \aleph_{2,Dim-1} & \aleph_{2,Dim} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \aleph_{v,c} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \aleph_{F-1,1} & \cdots & \aleph_{F-1,c} & \cdots & \aleph_{F-1,Dim-1} & \aleph_{F-1,Dim} \\ \aleph_{F,1} & \cdots & \aleph_{F,c} & \cdots & \aleph_{F,Dim-1} & \aleph_{F,Dim} \end{bmatrix} \tag{12}$$

Here, the location of the v^{th} person is indicated by $\aleph_v$. Equation (13) can be employed to compute each $\aleph_{v,c}$ in the matrix $\hbar$.

$$\aleph_{v,c} = y_1 \times \left(UB_c - LB_c\right) + LB,\ v = 1,2,...,F,\ c = 1,2,...,Dim \tag{13}$$

where F indicates total number of people in population, Dim indicates size of the problem's dimensions, and y_1 represents random number between 0 and 1. Moreover, UB_c *and* LB_c are upper and lower bounds of problem's c^{th} dimension.

Step ii): Fitness function

Since using an error function to obtain the best result is thought to be of minimal difficulty, the optimal solution is that which has the lowest Mean Square Error (MSE). Consequently, MSE is shown below.

$$MSE = \frac{1}{W}\sum_{\partial=1}^{W}\left[\varpi_h - \nu\right]^2 \tag{14}$$

Here, an expected output is represented as ϖ_h, categorized output from CNN is symbolized as ν, and the total number of samples is denoted as W.

Step iii): Exploration phase

Gannets alter their diving patterns to match the prey's dive depth while they search for prey in the water from the air. There are two distinct types of dives: a long, deep U-shaped dive and a short, shallow V-shaped dive.

$$\xi = 1 - \frac{Iter}{U_{\max\_iter}} \tag{15}$$

$$d = 2 \times \cos(2 \times \pi \times y_2) \times \xi \tag{16}$$

$$s = 2 \times \alpha(2 \times \pi \times y_3) \times \xi \tag{17}$$

$$\alpha(\aleph) = \begin{Bmatrix} -\frac{1}{\pi} * \aleph + 1, & \aleph \in (0, \pi) \\ \frac{1}{\pi} * \aleph - 1, & \aleph \in (\pi, 2\pi) \end{Bmatrix} \tag{18}$$

Here, $iter$ represents number of ongoing iterations, $U_{\max\_iter}$ represents number of iterations that can be performed, and y_2 *and* y_3 are random values between 0 and 1.

These two dive tactics will be used to update position in the following stage. In order to randomly select the two diving tactics when Gannet is predating, we define a random number ω. Gannet has roughly the same probability of selecting based on below equation (19).

$$\hbar_v(\xi + 1) = \hbar_v(\xi) + z_1 + z_2 \tag{19}$$

$$\text{Here, } z_2 = \kappa * (\hbar_v(\xi) - \hbar_y(\xi)) \tag{20}$$

$$\kappa = (2 \times y_4 - 1) \times d \tag{21}$$

Substitute equation (20) in equation (19), an expression becomes,

$$\hbar_v(\xi + 1) = \hbar_v(\xi) + z_1 + \kappa * (\hbar_v(\xi) - \hbar_y(\hbar)) \tag{22}$$

$$\hbar_v(\xi + 1) = \hbar_v(\xi) + z_1 + \kappa * (\hbar_v(\xi) - \kappa * \hbar_y(\hbar) \tag{23}$$

$$\hbar_v(\xi + 1) = \hbar_v(\xi)[1 + \kappa] + z_1 - \kappa * \hbar_y(\hbar) \tag{24}$$

From SSA [20], the update equation of location is expressed as,

$$\hbar_v(\xi + 1) = \hbar_{best}^{\xi} + \zeta . |\hbar_v(\xi) - \hbar_{best}^{\xi}| \tag{25}$$

$$\hbar_v(\xi + 1) = \hbar_{best}^{\xi} + \zeta . \hbar_v(\xi) - \zeta \hbar_{best}^{\xi} \tag{26}$$

$$\hbar_v(\xi+1)=\hbar_{best}^{\xi}[1-\zeta]+\zeta\hbar_v(\xi) \quad (27)$$

$$\zeta\hbar_v(\xi)=-\hbar_{best}^{\xi}[1-\zeta]+\hbar_v(\xi+1) \quad (28)$$

$$\hbar_v(\xi)=\frac{1}{\zeta}\left[\hbar_v(\xi+1)-\hbar_{best}^{\xi}(1-\zeta)\right] \quad (29)$$

$$\hbar_v(\xi+1)=\left(\frac{\hbar_v(\xi+1)-\hbar_{best}^{\xi}(1-\zeta)}{\zeta}\right)(1+\kappa)+z_1-\kappa*\hbar_y(\xi) \quad (30)$$

$$\frac{\hbar_v(\xi+1)\zeta-\hbar_v(\xi+1)(1+\kappa)}{\zeta}=\frac{\left(z_1-\kappa*\hbar_y(\xi)\right)\zeta-\hbar_{best}^{\xi}(1-\zeta)(1+\kappa)}{\zeta} \quad (31)$$

$$\hbar_v(\xi+1)\left[\zeta-1-\kappa\right]=\left(z_1-\kappa*\hbar_y(\xi)\right)\zeta-\hbar_{best}^{\xi}(1-\zeta)(1+\kappa) \quad (32)$$

$$\hbar_v(\xi+1)=\frac{1}{\zeta-1-\kappa}\left[\left(z_1-\kappa*\hbar_y(\xi)\right)\zeta-\hbar_{best}^{\xi}(1-\zeta)(1+\kappa)\right] \quad (33)$$

The best solution's position is updated using the equation above. Wherein, best solution is indicated as $\hbar_{best}^{\xi}$, step size control parameter is signified as ζ, random number $-d \text{ and } d$, randomly selected individual in ξ^{th} population is represented as $\hbar_y(\xi)$.

Step iv): Exploration phase

The exploitation process requires two further steps once the gannet rushes into the water in the manners mentioned above. Fish that are cunning in the water frequently make an abrupt turn in order to avoid the gannet's pursuit. In order to catch the fish desperately trying to get away, the gannet also uses a lot of energy. Here, the capture capacity is defined as,

$$\Delta=\frac{1}{A*\xi_2} \quad (34)$$

$$\xi_2=1+\frac{Iter}{U_{\max\_iter}} \quad (35)$$

$$A=\frac{G*velocity^2}{N} \quad (36)$$

$$N=0.2+(2-0.2)*y_6 \quad (37)$$

Here, capturability is represented as Δ, a random number between 0 and 1 is represented as y_6, the weight of gannet $G=2.5\,Kg$, and the gannet's water-moving speed is the $velocity=1.5m/s$. The next target at random using below equation (38).

$$G\hbar_v(\xi+1)=\begin{cases}\xi*delta*(\hbar_v(\xi)-\hbar_{best}(\xi))+\hbar_v(\xi), & \Delta\geq b\\ \hbar_{best}(\xi)-(\hbar_v(\xi)-\hbar_{best}(\xi))*\beta*\xi, & \Delta<b\end{cases} \quad (38)$$

Wherein, b represents constant value.

Step v): Re-evaluating the fitness

The position of gannet with the smallest objective function is considered optimal after changing its location, at which point equation (14), which recalculates the fitness of the solution, is used.

Step vi): Termination

The above steps are repeated until the pause condition is met. Algorithm 1 displays the GSSO method's pseudo code.

Hence, for the purpose of person recognition, the GSSO algorithm effectively determines the best person identification. The GSSO algorithm successfully obtains global optima with the least amount of difficulty by altering GOA in relation to SSA approach.

Algorithm 1. **Pseudo code of GSSO**

Input: Population size is F, problem dimension is Dim, maximum number of iterations is $U_{\max\_iter}$	
Output: The position of Gannet and value of its fitness	
1.	Initialize population $\hbar$, random numbers from 0 to 1 are $y \text{ and } \omega$
2.	Compute fitness value of $\hbar$
3.	**while** the ending condition is not satisfied **do**
4.	**if** rand>0.5 **then**
5.	Update location Gannet using Equation (19)
6.	**End if**
7.	**End for**
8.	**else**
9.	**If** $b\geq 0.2$
10.	Update location Gannet using Equation (33)
11.	**else**
12.	Update location Gannet using Equation (38)
13.	**End if**
14.	**End for**
15.	**End if**
16.	Analyze fitness value
17.	**End for**
18.	**End while**

4. Results and discussion

GSSO+DbneAlexNet attained best outcomes for person recognition using ear images and acquired results are elucidated in this section.

4.1 Experimental set-up

GSSO + DbneAlexNet designed in this paper for person recognition using ear images is carried out experimentally in the Python tool.

4.2 Dataset description

Here, 13 classes or groups of 2600 ear photos make up ear recognition Dataset's [17] ear images. This dataset was utilized to create the database for the final project and to train the suggested CNN deep neural network in this study. There are 520 photos in each of the test and validation sets, 1560 images in training set, and 008, 009, and 010 groups are used as the unknown groups in this dataset.

4.3 Experimental results

The experimental findings for the person recognition system GSSO + DbneAlexNet are reported in this section. In figure 4, the experimental results of person recognition using ear images. In figure 4a), b), c) indicates input image, pre-processed image, augmented image of flipping. Moreover, figure 4d), e), f) represents feature extracted image considering SIFT, feature extracted image considering PHoG, feature extracted image considering SURF.

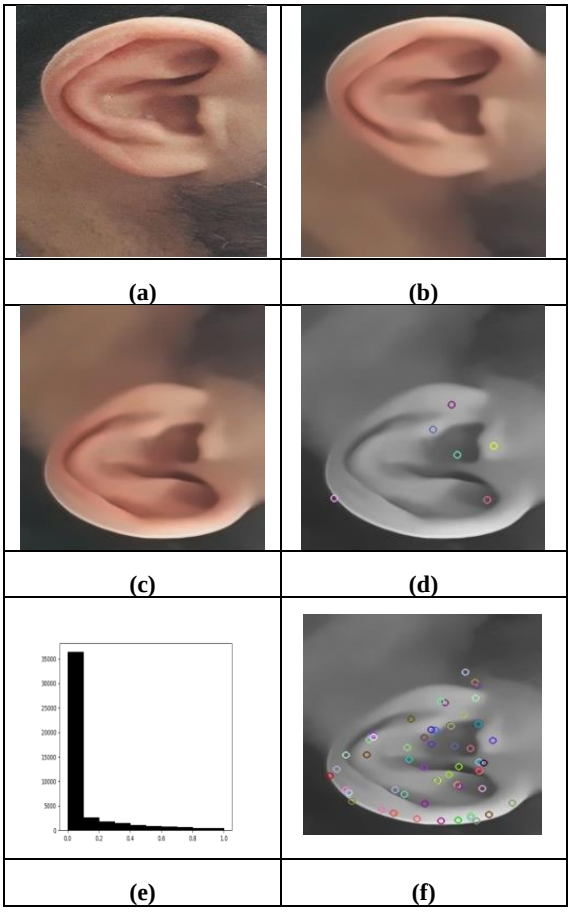

Figure 4. Experimental outcomes of GSSO+DbneAlexNet,

a) input image, b) pre-processed image, c) augmented image-flipping, d), e), f) indicates feature extracted image considering SIFT_flipping, feature extracted histogram for PHoG_flipping, feature extracted image considering SURF_flipping

4.4 Performance metrics

Precision, recall, and f-measure are the evaluation measures taken into account when making assessments.

i) Recall: It measures the proportion of true positives to sum of false negatives and true positives. It can be explained using following formula:

$$Z = \frac{K^{+ve}}{K^{+ve} + L^{-ve}} \tag{39}$$

Wherein, L^{-ve} represents false negative rate, Z signifies recall, and true positive rate is symbolized as K^{+ve}.

ii) Precision: By discriminating between true positives and false positives, the accuracy of actual sentiments is produced. In this scenario, both true and false positives are included in the total positives. The following is an example of precision:

$$\varphi = \frac{K^{+ve}}{K^{+ve} + L^{+ve}} \tag{40}$$

Where, K^{+ve} is the true positive rate, L^{+ve} indicates false positive rate and precision is indicated as φ.

iii) f-measure:Similar to the harmonic mean of recall and precision, which yields a weighted proportion and is written as,

$$\Omega = 2 \times \left(\frac{\varphi \times Z}{\varphi + Z} \right) \tag{41}$$

Here, Ω signifies f-measure, precision represents φ, and recall signifies Z.

4.5 Comparative techniques

DL [1], CNN [6], TL [7] and DCNN [4] are the techniques considered to compare with GSSO+DbneAlexNet to prove its efficiency.

4.6 Comparative analysis

An evaluation of GSSO+DbneAlexNet is implemented for ear images by consideration of performance metrics.

4.6.1 Assessment for person recognition using ear images

Precision, recall, and f-measure features are considered when training set differs as a feature of person recognition.

a) Analysis based on training set

The comparison analysis based on the training set is displayed in Figure 5. Figure 5(a) shows the GSSO+DbneAlexNet's precision.When considering training set=90% it appears that the conventional DL, CNN, TL, DCNN and proposed GSSO+DbneAlexNet achieved precision values of 82.5%, 83.5%, 86.5%, 88.6%, and 91.4%. This reveals that GSSO+DbneAlexNetoutperform other methods by 9.73%, 8.63%, 5.46%, and 3.06%.Recall serves as the basis for the investigation in Figure 3(b). The proposed GSSO+DbneAlexNet attained high recall value of 94.6% while current DL, CNN, TL, and DCNNhad low recall values of 86.5%, 87.7%, 90.7%, and 91.4% by considering training set=90%. Figure 3(c) shows the

comparative analysis of models for 90% training set based on F-measure. A high F-measure value of 93.0% was achieved by suggested GSSO+DbneAlexNet for training set=90%. When considering the training set=90%, the traditional approachesattainedf-measure values of 84.5% for the DL, 85.6% for the CNN, 88.6% for the TL, and 90.0% for theDCNN. The performance enhancementrevealed by GSSO+DbneAlexNet in this instance is 9.14%, 7.96%, 4.73%, and 3.23% better than existing methods like DL, CNN, TL, DCNN.

b) Analysis based on k-group

In this phase, the performance of GSSO+DbneAlexNet is evaluated based on K-group as exposed in figure 6. The precision of the GSSO+DbneAlexNet study is revealed in Figure 6a). The precision determined by GSSO+DbneAlexNet with K-group of 8 is 92.6%, which is higher by 9.61%, 7.34%, 5.29%, and 3.02% than the precision evaluated by methods like DL, CNN, TL, DCNNat 83.7%, 85.8%, 87.7%, and 89.8%.Figure 6b)displays assessment of GSSO+DbneAlexNetwith respect to recall. The recall'sofDLis87.7%, CNN is88.7%, TLis 91.6%, and DCNN is92.7% were all less than recall of 95.8% gained by the GSSO+DbneAlexNet with K-group of 8. This shows that the GSSO+DbneAlexNetproduce performance improvements of 8.46%, 7.41%, 8.56%, and 3.24%. The assessment of the GSSO+DbneAlexNet based on f-measure is portrayed in Figure 6c). The current systems, which include DL, CNN, TL, DCNN and GSSO+DbneAlexNet, have attained f-measure values of 85.6%, 87.2%, 89.6%, 91.2%, and 94.2%. This displays that the GSSO+DbneAlexNet yields improvementswith values 9.13%, 7.43%, 4.88%, and 3.18%correspondingly.

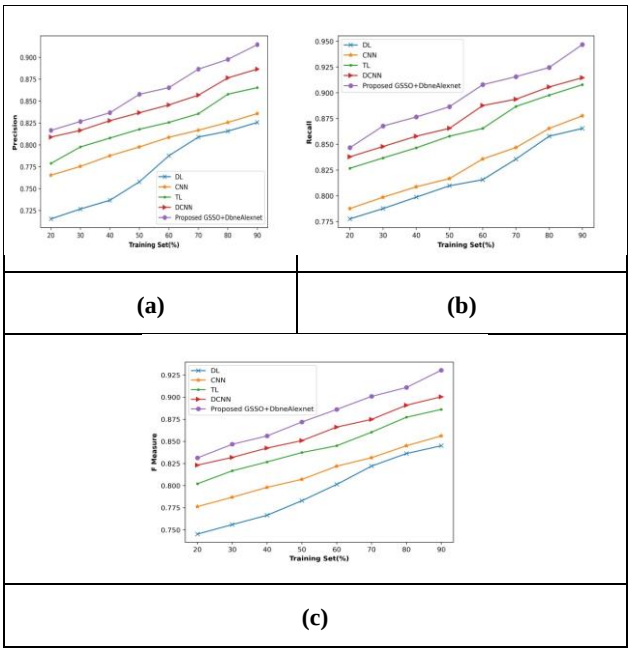

Figure 5. Comparative assessment in terms of training set,

a) precision, b) recall, c) f-measure

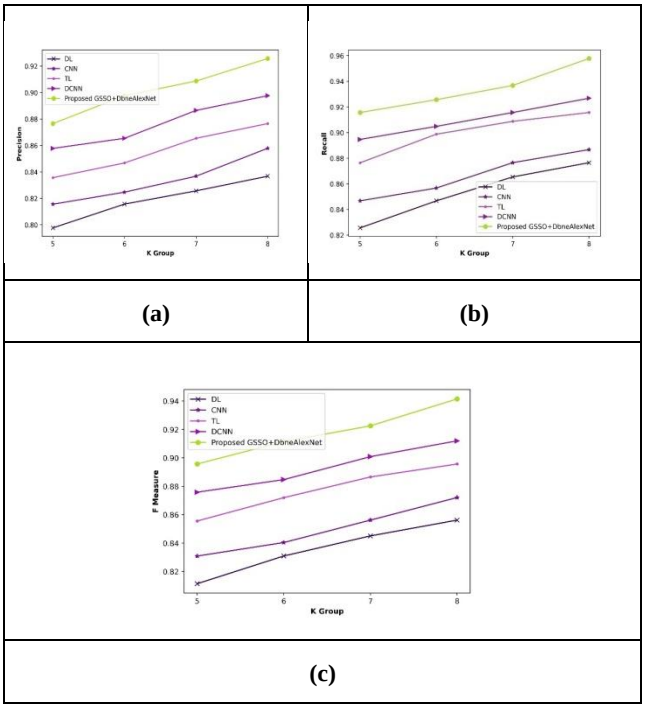

Figure 6. Comparative assessment based on k-group,

a) precision, b) recall, c) f-measure

4.7 Algorithmic techniques

GSSO_CNN-TL of algorithmic approaches which are utilized to evaluate, such as Crow Search Algorithm (CSA)+DbneAlexNet [26], Squirrel Search Algorithm (SSA)+DbneAlexNet [27], Sparrow Search +DbneAlexNet[28] and GOA+DbneAlexNet[19].

4.7.1 Algorithmic analysis

In figure 7, the proposed GSSO+DbneAlexNet's algorithm-based performance evaluation is shown. Changing the swarm size allows for evaluation in this case. Figure 7 displays the suggested GSSO+DbneAlexNet algorithm-based performance evaluation. In this scenario evaluation can be done by varying the swarm size. The assessment with respect to the precision parameter is shown in figure 7a). The swarm size of CSA+DbneAlexNet, SSA+DbneAlexNet, Sparrow Search+DbneAlexNet, GOA+DbneAlexNet and proposed GSSO+DbneAlexNet are 0.837, 0.858, 0.865, 0.887 and 0.915correspondingly. From figure 7 b), recall of developed GSSO+DbneAlexNet is 0.947 for swarm size=20. Here, other prevailingmethodsachievedrecall of CSA+DbneAlexNet is 0.848, SSA+DbneAlexNet is 0.865, Sparrow Search+DbneAlexNet is 0.887, and GOA+DbneAlexNet is 0.927. Furthermore, a valuationin terms of f-measure is revealed in figure 7 c). While swarm size=20, the f-measure of proposed GSSO+DbneAlexNet is 0.930, while f-measure of 0.842, 0.862, 0.876, and 0.906 are achieved by other prevailingapproaches like CSA+DbneAlexNet, SSA+DbneAlexNet, Sparrow Search+DbneAlexNet, and GOA+DbneAlexNet.

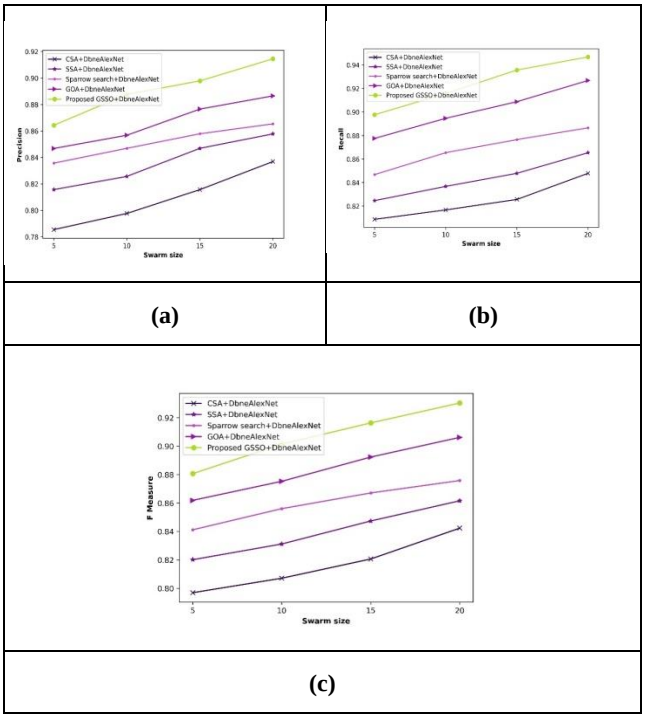

Figure 7. Algorithmic assessment of GSSO_CNN-TL in terms of a) precision, b) recall, c) f-measure

4.8 Performance analysis

Figure 8 shows the performance evaluation of the suggested GSSO+DbneAlexNet for person recognition utilizing multiple epochs. In this case, the performance of the GSSO+DbneAlexNet is assessed in terms of the training set over 20, 40, 60, and 80 epochs. Additionally, the performance is assessed based on the precision, recall, and f-measure characteristics. Moreover, figure 8 a) shows performance of GSSO+DbneAlexNet regards to precision. The suggested GSSO+DbneAlexNet provided precision of 89.8%, 90.8%, 91.5%, and 93.7% for 20, 40, 60, and 80 epochs for the training set=90%. The analysis related to recall is also considered in figure 8 b). The recall of the GSSO+DbneAlexNet was 92.6%, 93.7%, 94.4%, and 96.5% in the 20, 40, 60, and 80 epochs, respectively, whereas the training set recall was 90%. Figure 8(c) provides an illustration of the estimation with relation to the f-measure. The f-measure values 91.2%, 92.2%, 92.9%, and 95.1% are achieved in 20, 40, 60, and 80 epochs, for the training set=90%.

4.9 Comparative discussion

Table 1 describes comparison analysis of GSSO+DbneAlexNet. The outcomes for the proposed and current approaches for precision, recall, and f-measure are shown below. The 90% training set with K-group value of 9 yield highest values of these metrics.In training set-based assessment, the extreme value of precision 92.6% is attained by the proposed GSSO+DbneAlexNet, whereas the precisionachieved by the existing methods like DL, CNN, TL, and DCNN are 83.7%, 85.8%, 87.7%, and 89.8%. Furthermore, maximum recall value of 95.8% is attained by proposed GSSO+DbneAlexNet, while other approaches provided recall of 87.7%, 88.7%, 91.6%, and 92.7%. Likewise, the greatest f-measure of the suggested GSSO+DbneAlexNet is 94.2%, while the other approaches acquired f-measures of 85.6%, 87.2%, 89.6%, and 91.2% using other prevailing techniques. Likewise, maximum values of precision, recall, and f-measure are92.6%, 95.8%, and 94.2% are attained by proposed GSSO + DbneAlexNet using K-group variations. As a result, GSSO + DbneAlexNet training method performs better than the other methods discussed here.

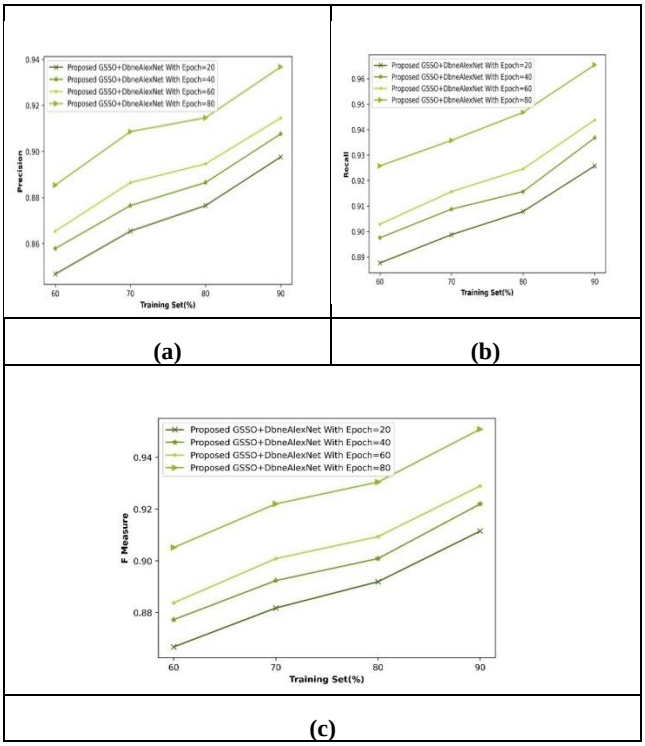

Figure 8. Performance analysis based on

a) precision, b) recall, c) f-measure

Table 1. Comparative discussion of proposed GSSO+DbneAlexNet

Person recognition	Metrics / Methods	DL	CNN	TL	DCNN	Proposed GSSO+DbneAlexNet
Based on training set=90	*Precision (%)*	82.5	83.5	86.5	88.6	91.4
	Recall (%)	86.5	87.7	90.7	91.4	94.6
	f-measure (%)	84.5	85.6	88.6	90.0	93.0
Based on K-group=8	*Precision (%)*	83.7	85.8	87.7	89.8	92.6
	Recall (%)	87.7	88.7	91.6	92.7	95.8
	f-measure (%)	85.6	87.2	89.6	91.2	94.2

5. Conclusion

Secure authentication is necessary for a number of purposes, including verifying a user's identity to access challenging systems, keeping an eye out for and identifying unauthorized people, and safeguarding assets from potential dangers. As a result, identifying people based on ear structure is preferable. In order to develop a powerful person recognition model, the proposed GSSO+DbneAlexNet is applied in this study. Initially, an input ear image is acquired from person recognition dataset and it is forwarded to the pre-processing phase. To remove noise from the input image during pre-processing, the NLM filter is used. Additionally, image rotation, flipping, cropping, and colour augmentation are all used in the image augmentation process.Once image is augmented, the features are extracted. The

features such as the SIFT, SURF, PHoG, and LBP are retrieved. Furthermore, the person recognition is done using CNN transfer learning where, CNN is used with hyper-parameters from the trained models like DbneAlexNet, which is trained using GSSO. By combining GOA with SSA in this instance, a novel GSSO has been developed. Precision, recall, and f-measure metrics for the k-group were 92.6%, 95.8%, and 94.2%, respectively, using the devised technique to achieve maximum performance. Future research will consider other deep learning classifier training optimization methodologies when determining the potential dimension of the probe.

6. References

[1] AhilaPriyadharshini, R., Arivazhagan, S. and Arun, M., "A deep learning approach for person identification using ear biometrics", Applied intelligence, vol. 51, pp.2161-2172, 2021.

[2] Hossain, S., Sultana Mitu, S., Afrin, S. and Akhter, S., "A Real-time Machine Learning-Based Person Recognition System with Ear Biometrics", International Journal of Computing and Digital System, 2021.

[3] Mursalin, M., Ahmed, M. and Haskell-Dowland, P., "Biometric Security: A Novel Ear Recognition Approach Using a 3D Morphable Ear Model", Sensors, vol. 22, no. 22, pp.8988, 2022.

[4] Alshazly, H., Linse, C., Barth, E. and Martinetz, T., "Deep convolutional neural networks for unconstrained ear recognition", IEEE Access, vol. 8, pp.170295-170310, 2020.

[5] Sarangi, P.P., Mishra, B.S.P., Dehuri, S. and Cho, S.B., "An evaluation of ear biometric system based on enhanced Jaya algorithm and SURF descriptors", Evolutionary Intelligence, vol. 13, pp.443-461, 2020.

[6] Sinha, H., Manekar, R., Sinha, Y. and Ajmera, P.K., "Convolutional neural network-based human identification using outer ear images", In Soft Computing for Problem Solving: SocProS 2017, Volume 2 (pp. 707-719), Springer Singapore, 2019.

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

[8] Yaman, D., Eyiokur, F.I. and Ekenel, H.K., "Ear2Face: Deep Biometric Modality Mapping", arXiv preprint arXiv:2006.01943, 2020.

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

[10] Sharkas, M., "Ear recognition with ensemble classifiers; A deep learning approach", Multimedia Tools and Applications, pp.1-27, 2022.

[11] Zhu, Q. and Mu, Z., "PointNet++ and three layers of features fusion for occlusion three-dimensional ear recognition based on one sample per person", Symmetry, vol. 12, no. 1, pp.78, 2020.

[12] Unar, J.A., Seng, W.C. and Abbasi, A., "A review of biometric technology along with trends and prospects", Pattern recognition, vol. 47, no. 8, pp.2673-2688, 2014.

[13] Wang, Z., Yang, J. and Zhu, Y., "Review of ear biometrics", Archives of Computational Methods in Engineering, vol. 28, pp.149-180, 2021.

[14] Ganapathi, I.I., Ali, S.S., Prakash, S., Vu, N.S. and Werghi, N., "A Survey of 3D Ear Recognition Techniques", ACM Computing Surveys, vol. 55, no. 10, pp.1-36, 2023.

[15] Zhao, W., Chellappa, R., Phillips, P.J. and Rosenfeld, A., "Face recognition: A literature survey", ACM computing surveys (CSUR), vol. 35, no. 4, pp.399-458, 2023.

[16] Minaee, S., Abdolrashidi, A., Su, H., Bennamoun, M. and Zhang, D., "Biometrics recognition using deep learning: A survey", Artificial Intelligence Review, pp.1-49, 2023.

[17] Datasets for ear detection and recognition, "https://www.kaggle.com/datasets/omarhatif/datasets-for-ear-detection-and-recognition", accessed on April, 2023.

[18] Kim, J.H., Lee, C., Sim, J.Y. and Kim, C.S., "Single-image deraining using an adaptive nonlocal means filter", In IEEE international conference on image processing, pp. 914-917, September 2013.

[19] Pan, J.S., Zhang, L.G., Wang, R.B., Snasel, V. and Chu, S.C., "Gannet optimization algorithm: A new metaheuristic algorithm for solving engineering optimization problems", Mathematics and Computers in Simulation, vol.202, pp.343-373, 2022.

[20] Xue, J. and Shen, B., "A novel swarm intelligence optimization approach: sparrow search algorithm", Systems science & control engineering, vol.8, no.1, pp.22-34, 2020.

[21] Khosla, C. and Saini, B.S., "Enhancing performance of deep learning models with different data augmentation techniques: A survey", In 2020 International Conference on Intelligent Engineering and Management (ICIEM), pp. 79-85, IEEE, June, 2020.

[22] Ghoualmi, L., Draa, A. and Chikhi, S., "An ear biometric system based on artificial bees and the scale invariant feature transform", Expert Systems with Applications, vol. 57, pp.49-61, 2016.

[23] Bai, Y., Guo, L., Jin, L. and Huang, Q., "A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition", In proceedings International Conference on Image Processing (ICIP), pp. 3305-3308, November, 2009.

[24] Lu, K., Luo, J., Zhong, Y. and Chai, X., "Identification of wool and cashmere SEM images based on SURF features", Journal of engineered fibers and fabrics, vol. 14, pp.1558925019866121, 2019.

[25] NarainPonraj, D., Christy, E., Aneesha, G., Susmitha, G. and Sharu, M., "Analysis of LBP and LOOP based textural feature extraction for the classification of CT Lung images", In 2018 4th International Conference on Devices, Circuits and Systems (ICDCS), pp. 309-312, IEEE, March, 2018.

[26] Askarzadeh, A., "A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm", Computers & structures, vol. 169, pp.1-12, 2016.

[27] Zheng, T. and Luo, W., "An improved squirrel search algorithm for optimization", Complexity, 2019.

[28] Ouyang, C., Zhu, D. and Wang, F., "A learning sparrow search algorithm", Computational intelligence and neuroscience, 2021.

[29] Yao, R., Wang, N., Liu, Z., Chen, P. and Sheng, X., "Intrusion detection system in the advanced metering infrastructure: a cross-layer feature-fusion CNN-LSTM-based approach", Sensors, vol. 21, no. 2, pp.626, 2021.

[30] Shi, Z., Hao, H., Zhao, M., Feng, Y., He, L., Wang, Y. and Suzuki, K., "A deep CNN based transfer learning method for false positive reduction", Multimedia Tools and Applications, vol. 78, pp.1017-1033, 2019.

[31] Alaeddine, H. and Jihene, M., "Deep Batch-normalized eLU AlexNet for Plant Diseases Classification", In 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD), pp. 17-22, IEEE, March, 2021.

[32] Imoto, K., Nakai, T., Ike, T., Haruki, K. and Sato, Y., "A CNN-based transfer learning method for defect classification in semiconductor manufacturing", IEEE Transactions on Semiconductor Manufacturing, vol. 32, no. 4, pp.455-459, 2019.