Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 4, 8182 – 8190

HUMAN EAR RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK

[1]T. Ebanesar, [2]A. D. Bibin, [3]J. Jalaja

[1]*Assistant Professor, Research Department of Computer Science, Malankara Catholic College, Mariagiri, Tamilnadu, India, ebanesarcs@gmail.com*
[2]*Assistant Professor, Research Department of Computer Science, Malankara Catholic College, Mariagiri, Tamilnadu, India*
[3]*Assistant Professor, Research Department of Computer Science, Malankara Catholic College, Mariagiri, Tamilnadu, India*

Abstract

Today's digital world, security plays an important role in everyday process in computer. The existing security levels are hacked by someone at anytime. A biometric based security system is expected to fulfil user's demand such as low error rates, high security levels, possibility of fake detection etc. This paper proposes an efficient ear based recognition technique using Convolutional Neural Network (CNN). It is non –intrusive methodology and attributes are probably the most common biometric feature used by humans to recognize one other. There are many advantages of using the ear as a source of data for human identification. An ear biometrics system consists of ear detection and ear recognition (authentication and identification) modules. In this paper, we used AMI (translated from Spanish as Mathematical. Analysis of Images) dataset. These images were captured using Nikon D100 camera. Once the pre-processing of image gets completed, the images are converted into one dimensional space. LBP method can be used for classification, recognition and detection of original image pixel further, histogram is used to obtain more discrimination features. In an existing work, researchers used K-Nearest Neighbours (KNN) method. It doesn’t work with large number of dataset and also it is difficult to calculate the distance in each dimension. The aim of the proposed system is to improve the security level with the help of Convolutional Neural Network (CNN). The proposed system gives the overall accuracy of 98.99 %.

Keywords: K-Nearest Neighbours Algorithm (KNN), Local Binary Pattern Method (LBP), Convolutional Neural Network (CNN), AMI Dataset, Ear Dataset, Ear Recognition, Feature Extraction

1. INTRODUCTION

A Convolutional Neural Network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. Convolutional Neural Network, also called ConvNet, was first introduced in the 1980s by Yann LeCun, a postdoctoral computer science researcher. The name "convolutional neural network" indicates that the network employs a mathematical operation called convolution. Today’s hacking world, we are struggling to secure our valuable data. For security process, we use the different number of methods to follow and keep the files in secure way. But, biometric objects are the best way to secure the data at highest level of security. CNNs are particularly useful for finding patterns in images to recognize objects, faces, and scenes. Machine learning offers us the tools and algorithms to analyze and process data to make accurate predictions. The world is filled with data—Images, videos, spreadsheets, audio, and text generated by people and computers are flooding the internet and drowning us in the sea of information. Figure 1 shows the detailed labelling of the human ear implored in the field of computer vision.

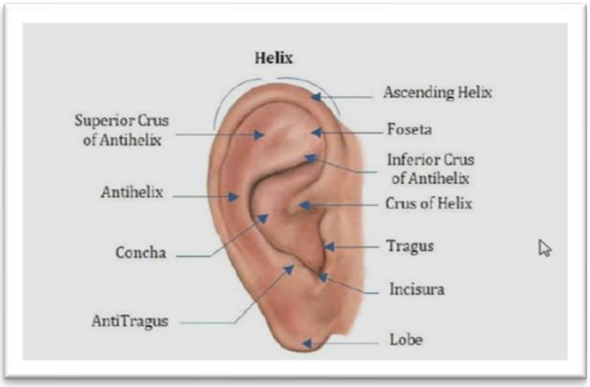

Figure 1: *The Structure of the Human Ear*

II. RELATED WORK

This section reviews on various techniques for object recognition which was done earlier that have been made in the area of face recognition. Anila S, Devarajan N. [1] [2020] proposed a method of preprocessing by combining some preprocessing algorithms like histogram equalization, gabber filter.

Raktim Ranjan Nath et al. [2] [2020] proposed a method for face detection using SVM (Support Vector Machine). Researchers using HOG (Histogram of Oriented Gradient) based face detector which gives more accurate results rather than other machine learning algorithms like Haar Cascade.

Jayanthi Raghavan and Majid Ahmadi [3] [2021] proposed a method for a modified CNN-based face recognition system. Accuracy rates of up to 96.2% is achieved using the proposed model in Extended Yale B

database. The proposed method performs better in comparison to all other approaches in Extended Yale B. In FERET database, CNN-LENET method performs better than the proposed method.

Syazana-Itqan K et al. [4][2020] proposed a method on developing a MATLAB-based Convolution Neural Network (CNN) face recognition system with Graphical User Interface (GUI) as the user input. The accuracy of the system is 100% for all 50 subjects.

Salman Mohammed Jiddah et al. [5] [2018] proposed a method for ear recognition using CNN. This research in respect to its aims and objectives carried out experiments on AMI ear database by extracting the Local Binary Patterns features and also applying Laplacian filter on the raw images separately to extract the geometric features. The accuracy of the system is 75.96%.

III. FRAMEWORK OF THE PROPOSED SYSTEM

The overall block diagram of the proposed human ear recognition system using convolutional neural network is shown in figure 2.

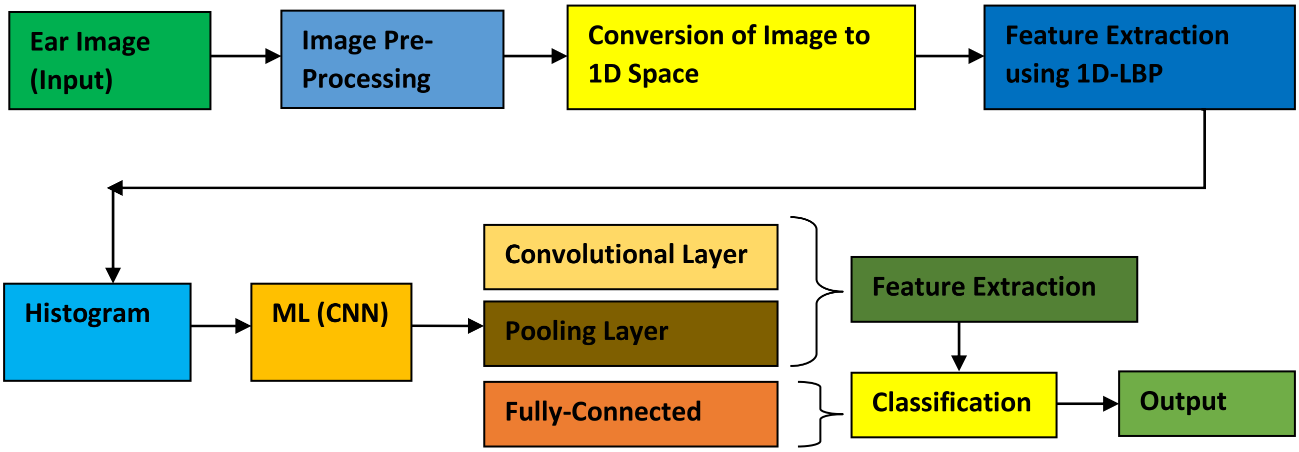

Figure 2: *The Proposed Framework of the Human Ear Recognition System*

1) IMAGE PRE-PROCESSING

Image pre-processing is considered as an important step in image recognition process. Today, most of the image data in the real world are incomplete, meaningless data and noisy data. The main aim of this image pre-processing is to remove the unnecessary noisy data which do not give useful information. Also it removes anomalous data from the dataset. There are four pre-processing steps applied in ear recognition system. It includes Clean Missing Data, Smooth Data, Fill Outliers and Extrema. The block diagram of an image pre-processing is shown in figure 3.

Data cleaning is the first step in any data processing pipeline. Messy data—heterogeneous values, missing entries, and large errors—is a major obstacle to automated modelling. Without clean data, our models will deliver misleading results and seriously harm our decision-making processes. Data cleaning not only refers to removing chunks of unnecessary data, but it's also often associated with fixing incorrect information within the dataset and reducing duplicates. It is used to identifying messy column-oriented data, cleaning multiple variables of data at a time, and iterating on and refining the cleaning process.

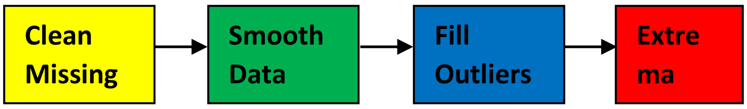

Figure 3: *Block Diagram of an Image Pre-Processing*

1.1) Clean Missing Data

Clean missing data process was done by using linear interpolation method. The interpolation is the process of finding out the unknown pixels of the image. There are various methods for interpolation. They are linear interpolation and polynomial interpolation. Image interpolation algorithms directly affect the quality of image magnification. Image interpolation is the process of transferring image from one resolution to another without losing image quality. Interpolation is used to estimate data points between two known points. The most common interpolation technique is Linear Interpolation. Interpolation - the process of finding a value between two points on a line or curve. Linear pattern - a pattern in which graphed points create a straight line. Linear interpolation is a method of curve fitting using linear polynomials to construct new data points within the range of a discrete set of known data points. The mathematical formula of Linear Interpolation is represented in the equation (1).

$$y = y_1 + (x - x_1)\,(y_2 - y_1)\; x_2 - x_1 \qquad \text{Equation----- (1)}$$

Unfortunately, missing data is unavoidable in poorly designed data collection procedures. It needs to be identified and dealt with as soon as possible. While these artifacts are easy to identify, filling up missing regions often requires careful consideration, as random fills can have unexpected outcomes on the model quality. Often, rows containing missing data are dropped as it's not worth the hassle to fill up a single data point accurately. When multiple data points have missing data for the same attributes, the entire column is dropped. Clean missing data of an image is shown in figure 4.

Figure 4: *Clean Missing Data of an Image*

Under completely unavoidable circumstances and in the face of low data, data scientists have to fill in missing data with calculated guesses. These calculations often require observation of two or more data points similar to the one under scrutiny and filling in an average value from these points in the missing regions.

1.2) Smooth Data

Data smoothing refers to techniques for eliminating unwanted noise or behaviors in data. In image processing, a Gaussian blur (also known as Gaussian smoothing) is the result of blurring an image by a Gaussian function (named after mathematician and scientist Carl Friedrich Gauss). It is a widely used effect in graphics software, typically to reduce image noise and reduce detail. In this image pre-processing Gaussian filter method was used to remove the noise or remove the grains. Data smoothing technique using Gaussian Filter is shown in figure 5.

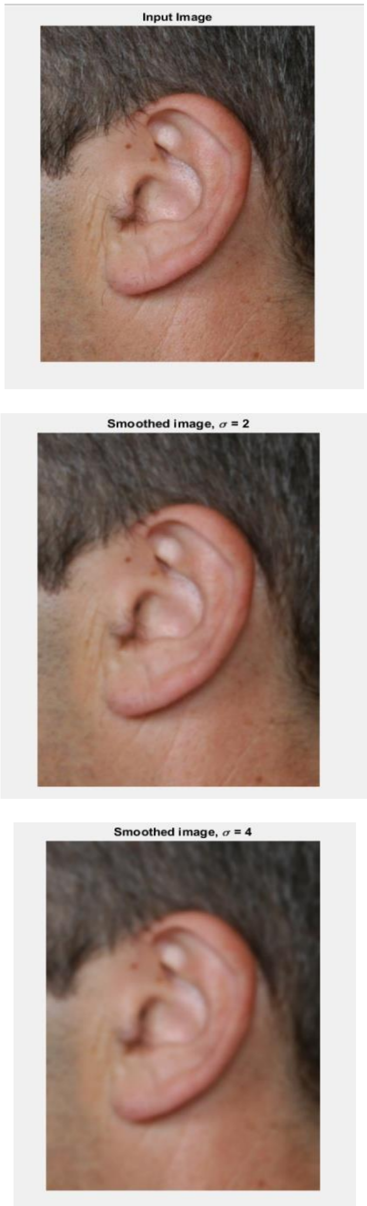

Figure 5: *Data Smoothing Technique Using Gaussian Filter*

1.3) Fill Outliers

Outlier detection is a primary task in image processing. Outlier detection identifies data points that are significantly different from the rest of the data. We can now remove the outliers from the cleaned data. Unwanted data in the form of outliers has to be removed before it can be processed further. Outliers are the hardest to detect amongst all other inaccuracies within the data. Thorough analysis is generally conducted before a data point or a set of data points can be rejected as an outlier. Specific models that have a very low outlier tolerance can be easily manipulated by a good number of outliers, therefore bringing down the prediction quality.

Low image data quality and the presence of noise bring a huge challenge to outlier detection. The raw image with an outlier will be incorporated and the image is converted into a black and white image then the pixels are extracted from the black and white image and its local likelihood values are calculated. Pixels of image were extracted and it is taken as 3 X 3 matrix format. Figure 6 represents the pixels present at the middle will be comparing with the nearby neighbor pixels values to detect the outliers presents in the image by finding the deviations. This process continues until each and every single pixels of the image get compared. The data values will be then categories as normal data and outlier data. Figure 7 describes the outlier of an image.

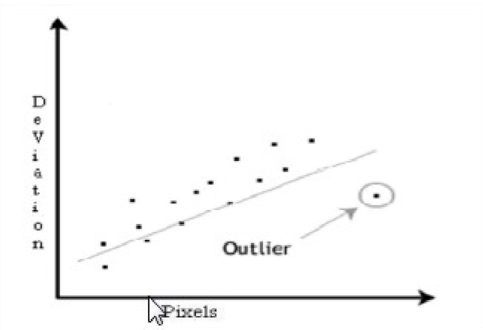

Figure 6: *Outlier Detection among the Image Pixels*

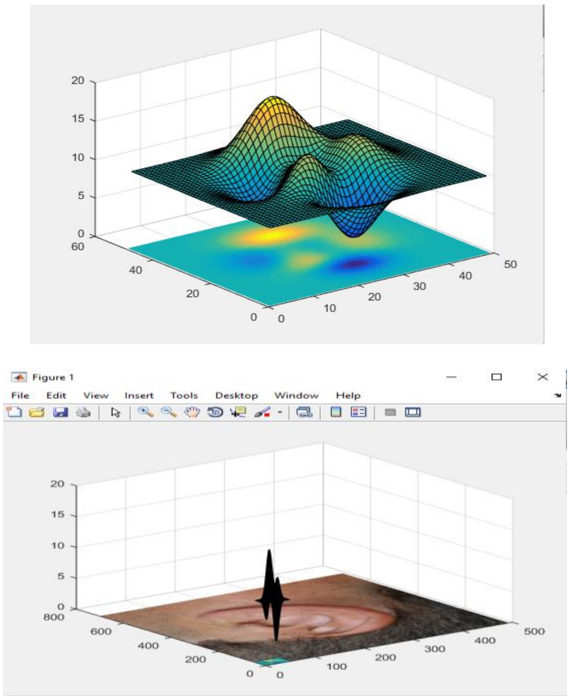

Figure 7: *Outlier of an Image*

1.4) Extrema

Finally, start typing the keyword extrema and click Find Local Extrema. Use smoothedData as the input data and change the extrema type to find both the local maxima and local minima of the cleaned, smoothed data. We can adjust the local extrema parameters to find more or fewer maxima and minima. Extrema of an image is shown in figure 8.

Figure 8: *Extrema of an Image*

2) CONVERSION OF IMAGE TO 1D SPACE

It is important to realize that images are stored as rectangular arrays of hundreds, thousands, or millions of discrete "picture elements," otherwise known as pixels. As noted, in practice, real world images will typically be made up of a vast number of pixels and each of these pixels will be one of potentially millions of colors. A grayscale image is an image in which a single pixel represents the amount of light or only contains light intensity information. It is a single-dimensional image and has different shades of gray color only. The color image 2D has RGB values with high intensity and huge number of pixels. During calculation time, it is very difficult to find the object. In order to avoid this, we convert the original 2D image into 1D image. In 2D image, for each pixel, take the average of the red, green, and blue pixel values to get the grayscale value.

As the grayscale images are single-dimensional, they are used to decrease models' training complexity in various problems. In MATLAB, an RGB image is basically an M*N*3 array of colour pixel, where each colour pixel is a triplet which corresponds to red, blue and green colour component of RGB image at a specified spatial location. Similarly, a grayscale image can be viewed as a single layered image. In MATLAB, a grayscale image is basically M*N array whose values have been scaled to represent intensities. Figure 9 shows the conversion of 2D image into 1D image.

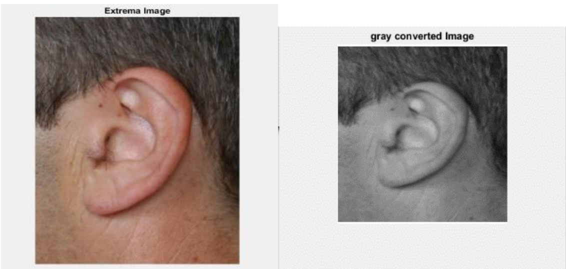

Figure 9: *Conversion of 2D image into 1D image*

3) FEATURE EXTRACTION USING 1D-LBP

LBP stands for Local Binary Pattern. One texture descriptor which has been used is Local Binary Pattern. (LBP). This is a method that depends on pixel-based texture extraction. With the development of local feature descriptors in other computer vision applications, the popularity of feature-based methods will be increased human ear recognition.

It is widely used in the applications based on image processing. The LBP works in a block size of 3×3 in which the center pixel is used as a threshold for the neighboring pixel and the LBP code of a center pixel is generated by encoding the computed threshold value into a decimal value. The mathematical expression of LBP is given in equation (2).

$$\mathrm{LBP} = \sum_{i=0}^{P-1} s(n_i - G_c)2^i$$

$$s(x) = \begin{cases} 1, & if\ x > 0 \\ 0, & otherwise \end{cases}$$

Equation---- (2)

where P is the number of neighborhood pixels, ni represents the ith neighboring pixel and c represents the center pixel. The histogram feature of size 2P is extracted from the obtained LBP code. Hence, for eight neighboring pixels the histogram feature vector length of 256 is obtained. The LBP operation is shown in figure 10 with a Gc value of 10 and eight neighboring pixels. Figure 11 shows the proposed LBP based human ear description.

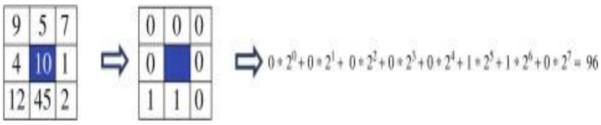

Figure 10: *LBP Operation with a Gc Value of 10 and Eight Neighboring Pixels*

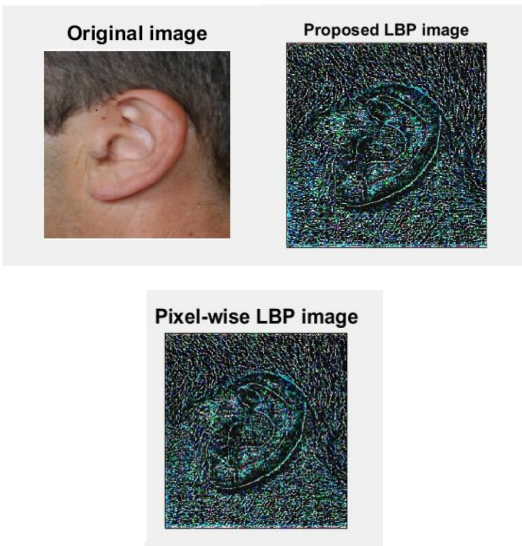

Figure 11: *LBP Based Human Ear Description*

4) HISTOGRAM

The most common pixel brightness transform operation is histogram. Histogram equalization is a well-known contrast enhancement technique due to its performance on almost all types of image. Histogram equalization provides a sophisticated method for modifying the dynamic range and contrast of an image by altering that image such that its intensity histogram has the desired shape. The formula of the normalized histogram is represented in equation (3). A very useful example of a gray level transform is histogram equalization. This transform flattens the gray level histogram of an image so that all intensities are as equally common as possible. This is often a good way to normalize image intensity before further processing and also a way to increase image contrast. The histogram equalization of an image is shown in figure 12.

P(n) = number of pixels with intensity n/ total number of pixels Equation-----(3)

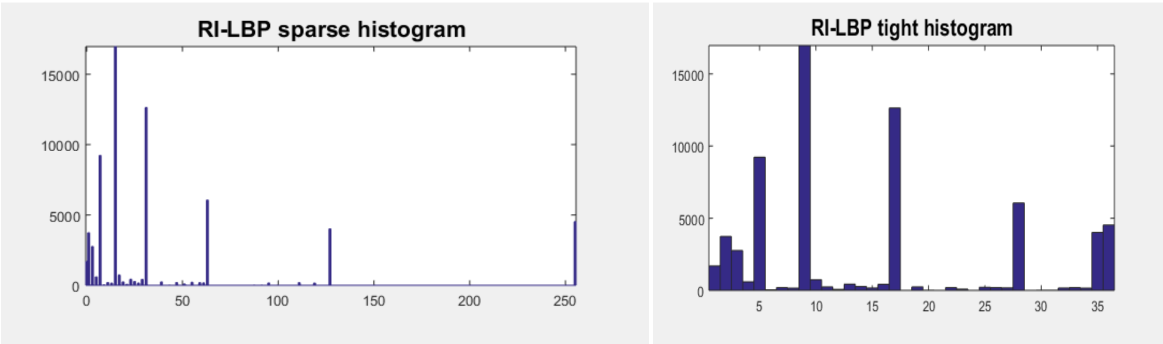

Figure 12: *Histogram Equalization of an Image*

5) CONVOLUTIONAL NEURAL NETWORKS

The Convolutional Neural Network (CNN) is a class of deep learning neural networks. A convolutional neural network is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. CNNs represent a huge breakthrough in image recognition. A CNN has Convolutional layers, ReLU layers, Pooling layers and a fully connected layer. CNNs can be used in tons of applications from image and video recognition, image classification, recommender systems to natural language processing and medical image analysis.

CNNs have an input layer, and output layer, and hidden layers. The hidden layers usually consist of convolutional layers, ReLU (Rectified Linear Unit) layers, pooling layers, and fully connected layers. Convolutional layers apply a convolution operation to the input. This passes the information on to the next layer. Pooling combines the outputs of clusters of neurons into a single neuron in the next layer. Fully connected layers connect every neuron in one layer to every neuron in the next layer. A CNN works by extracting features from images. This eliminates the need for manual feature extraction. The features are not trained. They're learned while the network trains on a set of images. This makes deep learning models extremely accurate for computer vision tasks. CNNs learn feature detection through tens or hundreds of hidden layers. Each layer increases the complexity of the learned features. Figure 13 illustrates the end-to-end structure of a convolutional neural network.

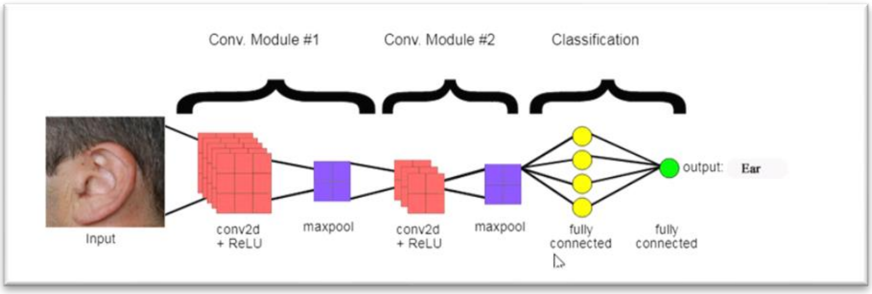

Figure 13: *End-to-End Structure of a Convolutional Neural Network*

IV. DATASET

Dataset is the foundation for any machine learning and neural network.Building great neural network model requires a huge amount of high quality labelled training data. In fact, eighty percent of the time spent on developing neural network is related to data set. In our project, we used AMI dataset for training. In building artificial neural network applications, datasets are divided into two subsets: i) Training data ii) Testing data. In a supervised learning, we use a training dataset that contains outcomes, to train the machine. Then we use testing dataset that has no outcomes to predict outcomes. The sample dataset of an ear image is shown in figure 14.

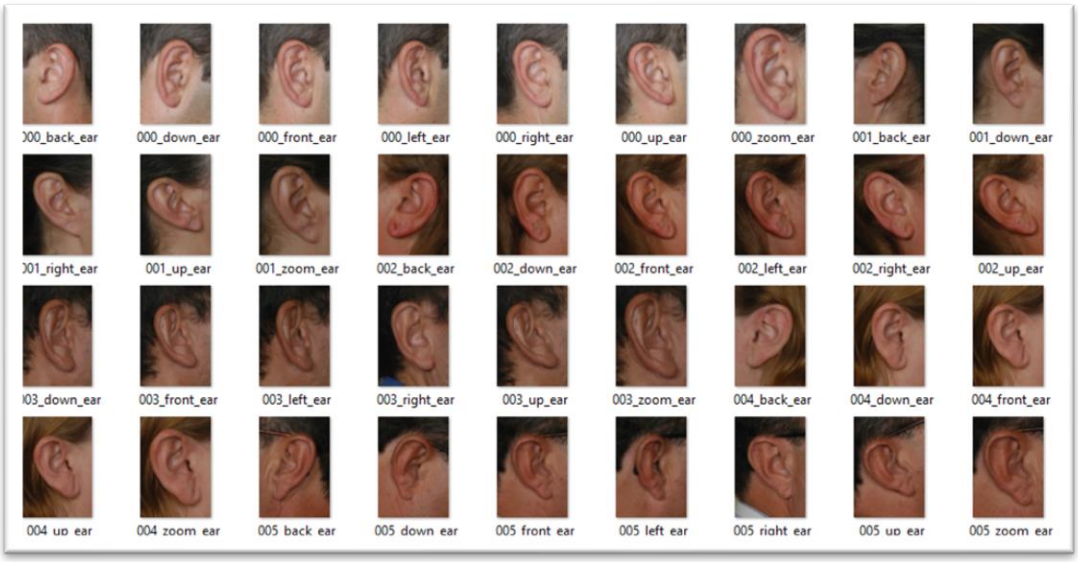

Figure 14: *AMI Ear Dataset Images*

In our research, researcher used the AMI ear database to carry out experiments. This database contains 600 images in total; of 100 subjects ranging from 19 years to 65 years of age, all images are in JPEG format with 492 X 702 pixels each. Each subject has seven images; 1 left ear image and six right ear images with varying positioning taken under the same lighting conditions. All experiments carried out were done on the right ear images which reduced our database to 600 ear images.

V. EXPERIMENTAL RESULT AND DISCUSSIONS

The experiment was carried out on 600 human ear images. From each image, the texture based features are extracted. Implementation of this work has been done using MATLAB. The performance of the LBP method is shown in table 1.

Table 1: *LBP Method Performance*

Iteration(s)	Training Images	Testing Images	Accuracy (%)
1	500	100	90
2	500	100	89
3	500	100	91
4	500	100	88
5	500	100	92
Ave. Accuracy			90

Table 1 shows the performance of the LBP ear images database performance which recorded an average of 90% accuracy after cross validation. The accuracy of the CNN is shown in table 2.

Table 2: *The Accuracy of the CNN*

Iteration(s)	Training Images	Testing Images	Accuracy (%)
1	500	100	98.9
2	500	100	99
3	500	100	99
4	500	100	98.9
5	500	100	98.99
Ave. Accuracy			98.99

Table 2 shows the performance of the CNN ear images database performance which recorded an average of 98.99% accuracy after cross validation.

VI. CONCLUSION

This paper proposes a work on human ear recognition system based on CNN neural network algorithm. The proposed human ear recognition system achieves 98.99% accuracy.

Reference

[1] Anila S, Devarajan N,"Pre-Processing Technique for Face Recognition Applications Under Varying Illumination Conditions", Global Journal of Computer Science and Technology July 2020.

[2] Raktim Ranjan Nath,Kaberi Kakoty and Dibya Jyoti Bora (2020), "Face Detection and Recognition Using Machine Learning Algorithm", Sambodhi (UGC Care Journal) ISSN: 2249-6661 Vol-43, No.-03 (III) November-December (2020).

[3] Jayanthi Raghavan and Majid Ahmadi (2021),"A Modified CNN-Based Face Recognition System", International Journal of Artificial Intelligence and Applications (IJAIA), Vol.12, No.2, March 2021.

[4] Syazana-Itqan K, Syafeeza A.R and Saad N.M (2020),"A MATLAB-Based Convolutional Neural Network Approach for Face Recognition System", Journal of Bioinformatics and Proteomics Review,Volume 2: Issue 1,January-2020.

[5] Salman Mohammed Jiddah and Kamil Yurtkan(2018),"Fusion of Geometric and Texture Features For EarRecognition" 2018.