# Face Recognition Using ELM with ResNet50

Robins Anand[1], and Tripti Goel[1]

[1] Department of Electronics and Communication, National Institute of Technology Silchar, Assam, India

[2] robinsanand222@gmail.com, triptigoel@ece.nits.ac.in

**Abstract.** Facial recognition is a complex problem that has received a great deal of attention due to its numerous applications, including security, surveillance, and identification. In this study, we provide a novel method that combines the strength of the Extreme Learning Machine (ELM) algorithm with the ResNet50 deep neural network for accurate and efficient face recognition. The two primary steps in our method are feature extraction and categorization. In the first step, we use the ResNet50 network to extract complex features from facial photos. An ELM classifier is then fed these features, which is a fast and efficient learning algorithm that is particularly suited for high-dimensional data. To determine whether our strategy is effective, we conducted experiments on a popular face recognition dataset, namely, *AT&T* Dataset on Faces. According to the results of our experiments, our suggested strategy is more accurate and effective than a number of cutting-edge approaches. Our findings demonstrate that the combination of ResNet50 and ELM gives a strong and effective response to the face recognition issue. This approach has significant potential for real-world applications where high accuracy and speed are crucial.

**Keywords:** Face recognition, Extreme Learning Machine, ResNet50, Feature extraction, Classification

## 1 Introduction

Face recognition (FR) technology, which identifies people based on their features, is widely used in several spheres of daily life. Features extraction and classification are the two major components of a face recognition system. The extracted features' efficiency significantly impacts how well a face recognition system performs [9]. Law enforcement, matching of photos on personal documents, customer authentication in financial transactions, access permission to specific databases or networks like those in government and business, and security screening at airports to prevent terrorist attacks are all current fields of appliances.

One of the most challenging issues for face recognition is lighting variance, which has drawn much attention recently. It is common knowledge that image variance caused by variations in lighting is more significant than variation caused



R. Murugan et al. (eds.), *Proceedings of the International Conference on Signal Processing and Computer Vision (SIPCOV 2023)*, Advances in Engineering Research 239,
https://doi.org/10.2991/978-94-6463-529-4_13

by differences in a personś identity [13]. However, uncontrolled situations continue to provide a number of difficulties. Some of these difficulties are brought on by issues with differences in lighting, face posture, expression, and other factors. One of the difficult issues that a realistic face recognition system must deal with is the impact of fluctuation in illumination conditions, which produces drastic variations in facial appearance.

In recent years, machine learning-based approaches have shown significant promise in the field of face recognition. In particular, Convolutional Neural Network (CNN) [3] serves as a tool for picture recognition processing of pixel data and extracting distinguishing face features. But now, the ResNet 50 [8] deep neural network has emerged as an accurate and efficient solution for face recognition. ResNet 50 is a neural network that is pre-trained on a large-scale dataset and is capable of extracting high-level features from facial images.

In order to categorize images, Softmax is a commonly used classifier in face recognition. For multi-class categorization, this activation function is utilized. But softmax models are prone to overfitting, especially when the number of parameters is high. These models are sensitive to the distribution of the training data. If the training data is imbalanced, the softmax model may produce biased predictions. Unlike softmax, ELM [8] is less sensitive to the distribution of the data, as it randomly initializes the weights in the hidden layer, which helps to prevent bias and it is less prone to overfitting, as it has a simple and efficient regularization mechanism. By combining ResNet50 and ELM, it is possible to extract discriminative features from facial images and classify them accurately and efficiently.

The proposed approach has several advantages over existing methods, including high accuracy, fast processing speed, and robustness to variations in facial expressions, poses, and lighting conditions. In this study, we provide a detailed methodology for facial identification using the ResNet50 with ELM approach and evaluate its effectiveness on popular face recognition datasets.

The following are the primary contributions of the suggested work:

1. The data is fed to the feature extractor, ResNet50.
2. ELM is used for the classification of images.
3. Computed the average classification accuracy.

The remaining sections are arranged as follows: In 2nd section, we present an extensive analysis of the available literature. on face recognition. In 3rd section, we describe the proposed methodology, including dataset preparation, feature extraction, dimension reduction, ELM training, and evaluation. Section 4 presents the experimental results and compares the proposed approach with other state-of-the-art methods. Section 5 concludes the paper by going over the potential applications and implications of the proposed system in the area of facial recognition.

## 2 Literature Review

Face recognition is an important research area that has received a lot of attention recently. It's a biometric technology that uses the unique physical features of an individual's face, such as the spacing between the eyes and the nose's shape, and the contours of the jawline, to identify them [15]. Face recognition systems typically work by capturing an image of a face using a camera or a video stream, and then processing that image using machine learning algorithms to extract and compare the facial features against a database of known faces [12].

Face recognition using machine learning algorithms has been common research. Many approaches have been put forth for this objective. Data in pixels is reduced in dimensions using Principal Component Analysis (PCA) [6]. It helps to find patterns in high-dimensional datasets and resize the images. Latent Dirichlet Allocation (LDA) [1] is used before the classification procedure to trim down the number of features to a more manageable quantity. It can be used as the preprocessing steps for the models. SVM [5] is a particular kind of deep learning algorithm. that carries out supervised learning for data group classification or regression. The algorithm draws a line or a hyperplane to divide the data into several classes. Convolutional Neural Networks (CNNs) [14] are used for feature extraction and categorization. The purpose of CNNs is to automatically recognize and extract features from photos. through multiple layers of convolution and pooling operations.

Kaur et. al [6] proposed this model, which effectively recognizes faces using the Principal Component Analysis(PCA) algorithm. This algorithm can recognize faces with an accuracy of 93.7% or higher.

Bansal et. al [1] proposed a model for facial recognition using PCA and LDA.

Geetha et. al [4] proposed a face identification and recognition system that uses the SVM machine learning algorithm to keep an eye on students during online tests.

Guo et. al [5] proposed an SVM-based facial recognition algorithm on the Cambridge facial database ORL. The nearest center classification (NCC) criterion was used to contrast the traditional eigenface technique with the SVM-based recognition.

Coskun et. al [3] suggested altering the CNN architecture by including two normalization procedures in two of the face recognition layers.

In this research paper, we are using deep learning for face recognition. A form of machine learning known as "deep learning" includes training artificial neural networks with numerous layers of processing units to recognize complex patterns and correlations in vast amounts of data. [7]. Deep learning models are capable of learning and extracting features automatically from raw data without requiring manual feature engineering. The accuracy and resilience of face recognition systems have been enhanced by the widespread usage of deep learning techniques. [11].

In this proposed work, we have used ELM for image classification. A single hidden layer with randomly initialized weights and a linear output layer make up the feedforward neural network known as ELM. Unlike traditional neural networks, ELM can be trained very efficiently, as the weights in the hidden layer are not changed throughout training.

The motivation for research on face recognition using ELMs stems from the need for efficient and accurate security systems, the advancements in technology, the limitations of existing methods, and the growing demand for face recognition. ELMs are a relatively new approach to machine learning, and there is still much to be learned about their capabilities and limitations. By exploring novel techniques for face recognition using ELMs, we could contribute to the development of this field.

## 3 Methodology

The dataset for this research paper is taken from Kaggle named as *AT&T* dataset of faces. The dataset has 400 pictures with 40 different subjects. The dataset is having 10 images of each person. Then the dataset is fed for extracting features. High-level characteristics are extracted from the facial photos using ResNet50 [8]. At last, ELM is used as a classifier on the reduced feature vectors.

### 3.1 Dataset

The Dataset used for this research work is *AT&T* Database of Faces (https://www.kaggle.com/datasets/kasikrit/att-database-of-faces) containing 400 grayscale pictures of 40 people. Each person has ten views. All *AT&T* database photographs were captured against a uniformly dark background with rotational and scaling changes up to 20. Several participants were photographed at different points in time with varying lighting, facial expressions (glances, closed or open eyes), and facial attributes (spectacles or no spectacles).

**Fig. 1.** Face pictures of the same individual from the AT&T database

The images are 92 x 112 pixels in size and have 256 different shades of grey per pixel. The files are in the PGM format.

### 3.2 Feature Extraction using ResNet50

ResNet 50 is a deep neural network architecture that has been trained beforehand on a sizable dataset [8]. For a number of computer vision tasks, such as object identification, it has shown higher performance. The feature extraction procedure for face recognition using ResNet 50 is described in this section.

ResNet 50 consists of 50 layers, comprising convolutional, pooling, and fully linked layers. Each layer of the model is trained to retrieve essential information from the input image. The final layer of ResNet 50 is a fully connected layer that outputs a 2048-dimensional feature vector, which represents the input image [14]. We use the final layer's output as the feature vector for the facial image. This feature vector stores detailed information about facial characteristics, including the appearance of the eyes, nose, and mouth, and the texture of the skin.

### 3.3 Classification using ELM

Machine learning algorithms like ELM are frequently employed for categorization jobs [10]. Face recognition is a popular application of ELM, where the algorithm is skilled in face image training and then put to use to categorize new face photos in accordance with how closely they resemble the training data.

Here's how the ELM algorithm has been trained using the training set for face recognition. Let $X$ be the input matrix with $n$ rows and $m$ columns, where n is the number of samples, m is the number of features and the last column represents the labels. Let $H$ be the hidden layer output matrix and $Y$ be the output matrix with $n$ rows and 1 column, where each element represents the predicted label for the corresponding input sample [2].

Firstly, initialized the weight matrix $W$ of the hidden layer randomly, with dimensions $(m-1)$ x $k$, where $(m-1)$ is the number of input features (excluding the label column). Compute the output of the hidden layer using the formula,

$$H = G(XW + b) \tag{1}$$

where $G$ represents activation function which is $Sigmoid$ in this paper, $H$ is a matrix with $n$ rows and $k$ columns and b is the bias.

Then, the output weight matrix, $\beta$ is computed using formula,

$$\beta = H^{-1}Y \tag{2}$$

and the predicted output matrix $Y_s$ is computed using,

$$Y_s = H * \beta \tag{3}$$

where H is the output of the hidden layer computed in equation 1, and $\beta$ is the output weight matrix computed in equation 2.

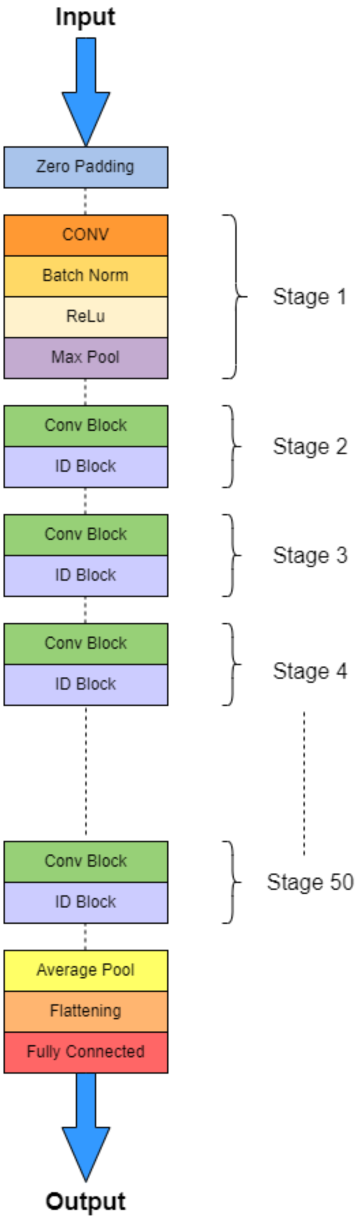


**Fig. 2.** Block diagram of ResNet 50

Classification error, $E$ is computed using the formula,

$$E = \frac{1}{2} * (Y - Y_s)' * (Y - Y_s)$$

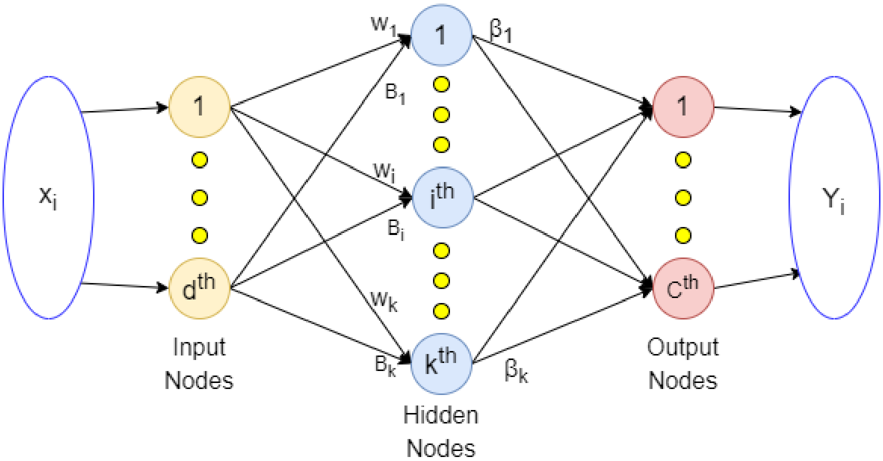


**Fig. 3.** Architecture of ELM network

Finally, the ELM algorithm has been tested using a testing set of values by feeding the feature vectors from the testing set to ELM and comparing the predicted outputs with the true outputs.

## 4 RESULTS and DISCUSSION

### 4.1 Implementation Details

In this research paper, the AT&T database is used with 400 images having resolution 0f 92 x 112 pixels. The number of images per subject is 10. There are 40 different subjects. After collecting the data, data was partitioned into training and testing sets. ResNet 50 has been used for extraction and ELM for the classification of images.

This project has been performed in MATLAB 2011 software, in a machine of 64bit operating system, 8 GB RAM, Windows 11 with the processor of Intel(R) Core i5- 10300 CPU at 2.50 GHz.

### 4.2 Comparison on changing the percentage of image per subject

In this research work, different split percentages have been taken i.e. different percentages of image per subject training, keeping other factors constant and checking the average testing accuracy and the change in split percentage affects the average testing accuracy. Here, we have kept the number of neurons as 1000 and a sigmoid activation function is used.

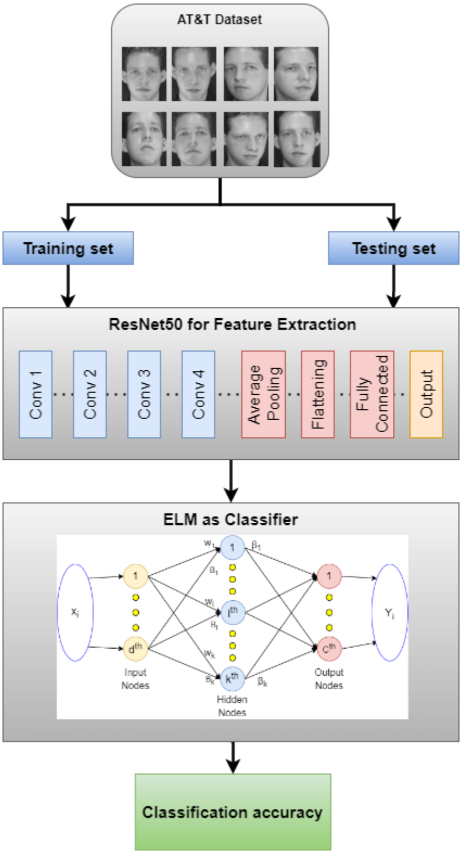


**Fig. 4.** Block diagram of proposed ELM algorithm

### 4.3 Comparison with different numbers of neurons in ELM classifier

In this experiment, different numbers of neurons are taken i.e. we keep changing the number of neurons, keeping other factors constant, and checking the average testing accuracy and how the change in neuron number affects the average testing accuracy.

### 4.4 Comparison with different activation functions

Activation functions are necessary for neural networks since they provide non-linearity to the network. Because of their non-linearity, neural networks can recognize intricate representations and functions based on their inputs which would not be possible with a simple linear regression model. In this experiment,

**Table 1.** Comparison on changing the percentage of image per subject

| **No. of neurons = 1000** | **Activation Function= Sigmoid** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Split percentage | Testing accuracy | Accuracy | Sensitivity | Specificity | Precision | Recall | F1 score | G-mean |
| 10% | 0.8433 | 0.869444 | 0.5555 | 0.877493 | 0.104167 | 0.555556 | 0.175439 | 0.698209 |
| 20% | 0.9568 | 0.94375 | 1 | 0.942308 | 0.307692 | 1 | 0.470588 | 0.970725 |
| 30% | 0.9702 | 0.971429 | 1 | 0.970696 | 0.466667 | 1 | 0.636364 | 0.985239 |
| 40% | 0.9714 | 0.995833 | 1 | 0.995726 | 0.857143 | 1 | 0.923077 | 0.997861 |
| 50% | 0.9828 | 0.985 | 0.8 | 0.989744 | 0.666667 | 0.8 | 0.727273 | 0.889829 |
| 60% | 0.985 | 0.9875 | 1 | 0.987179 | 0.666667 | 1 | 0.8 | 0.993569 |
| 70% | 0.9832 | 0.958333 | 1 | 0.957265 | 0.375 | 1 | 0.545455 | 0.978399 |
| **80%** | **0.9947** | **1** | **1** | **1** | **1** | **1** | **1** | **1** |
| 90% | 0.9835 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**Table 2.** Comparison with different numbers of neurons in ELM classifier

| Number of neurons | Testing accuracy | Accuracy | Sensitivity | Specificity | Precision | Recall | F1 score | G-mean |
|---|---|---|---|---|---|---|---|---|
| 100 | 0.7337 | 0.65 | 0.5 | 0.653846 | 0.035714 | 0.5 | 0.066667 | 0.571772 |
| 200 | 0.7625 | 0.65 | 0 | 0.807692 | 0 | 0 | | 0 |
| 300 | 0.6203 | 0.7125 | 0 | 0.730769 | 0 | 0 | | 0 |
| 400 | 0.103 | 0.1875 | 0 | 0.192308 | 0 | 0 | | 0 |
| 500 | 0.4587 | 0.5125 | 0 | 0.525641 | 0 | 0 | | 0 |
| 600 | 0.857 | 0.9 | 0.5 | 0.910256 | 0.125 | 0.5 | 0.2 | 0.674632 |
| 700 | 0.9487 | 0.9625 | 1 | 0.961538 | 0.4 | 1 | 0.571429 | 0.980581 |
| 800 | 0.979 | 0.9875 | 1 | 0.987179 | 0.666667 | 1 | 0.8 | 0.993569 |
| 900 | 0.9908 | 0.9875 | 1 | 0.987179 | 0.666667 | 1 | 0.8 | 0.993569 |
| **1000** | **0.9947** | **1** | **1** | **1** | **1** | **1** | **1** | **1** |
| 1100 | 0.991 | 0.9875 | 1 | 0.987179 | 0.666667 | 1 | 0.8 | 0.993569 |
| 1200 | 0.992 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

different activation functions are used, keeping other factors constant and checking how the average testing accuracy varies. Here, the number of neurons is kept as 1000.

**Table 3.** Comparison with different activation functions

| Activation Function | Testing accuracy | Accuracy | Sensitivity | Specificity | Precision | Recall | F1 score | G-mean |
|---|---|---|---|---|---|---|---|---|
| **Sigmoid** | **0.994** | **1** | **1** | **1** | **1** | **1** | **1** | **1** |
| Sine | 0.0277 | 0.025 | 0 | 0.025641 | 0 | 0 | | 0 |
| Hardlimit | 0.985 | 0.9875 | 1 | 0.987179 | 0.666667 | 1 | 0.8 | 0.993569 |
| tribas | 0.1743 | 0.2 | 0 | 0.205128 | 0 | 0 | | 0 |
| radbas | 0.3085 | 0.3375 | 0.5 | 0.333333 | 0.018868 | 0.5 | 0.036364 | 0.408248 |

### 4.5 Discussion

In this research paper, the number of images taken is 400 and 40 for different subjects. Each subject is having 10 images with a resolution of 92 x 112 pixels. The maximum accuracy on changing the split percentage while keeping other factors constant is 0.9947 on 80% split percentage for training as shown in table 1.The best average testing accuracy is recorded as 0.9947 on 1000 neurons after comparing average accuracy while varying the number of neurons. The effective activation function for this research is Sigmoid. The main advantage to use this activation function is that it exists between (0 to 1). It gives a smooth gradient.

## 5 Conclusion

In this study, we suggest a novel and efficient method for feature extraction and classification using ResNet 50 and ELM respectively. As a result, both ELM and feature extraction using ResNet 50 are effective approaches for face recognition tasks. ELM is a fast and simple algorithm that can be trained quickly on large datasets of face images. It can also handle high-dimensional feature spaces and is less prone to over-fitting, making it suitable for real-time face recognition applications, while feature extraction using ResNet 50 is a more sophisticated method that can extract high-level features from face images. By extracting features from the intermediate layers of ResNet 50, we can obtain a more compact and informative representation of the face images, which can be used for classification tasks. It yields accuracy rates of 0.9947 with ten images per person on the AT&T Laboratories Cambridge database. Without applying any preprocessing to the images from these databases, these results were achieved. It has a lower error rate than other available techniques. As a result, the suggested method recognizes faces for use in real-time applications with the highest degree of accuracy in the least period of time.

## References

1. Bansal, A., Mehta, K., Arora, S.: Face recognition using pca and lda algorithm. In: 2012 second international conference on Advanced Computing & Communication Technologies. pp. 251–254. IEEE (2012)
2. Boussaad, L., Boucetta, A.: Extreme learning machine-based age-invariant face recognition with deep convolutional descriptors. International Journal of Applied Metaheuristic Computing (IJAMC) 13(1), 1–18 (2022)
3. Coşkun, M., Uçar, A., Yildirim, Ö., Demir, Y.: Face recognition based on convolutional neural network. In: 2017 International Conference on Modern Electrical and Energy Systems (MEES). pp. 376–379. IEEE (2017)
4. Geetha, M., Latha, R., Nivetha, S., Hariprasath, S., Gowtham, S., Deepak, C.: Design of face detection and recognition system to monitor students during online examinations using machine learning algorithms. In: 2021 international conference on computer communication and informatics (ICCCI). pp. 1–4. IEEE (2021)
5. Guo, G., Li, S.Z., Chan, K.: Face recognition by support vector machines. In: Proceedings fourth IEEE international conference on automatic face and gesture recognition (cat. no. PR00580). pp. 196–201. IEEE (2000)
6. Kaur, R., Himanshi, E.: Face recognition using principal component analysis. In: 2015 IEEE international advance computing conference (IACC). pp. 585–589. IEEE (2015)
7. Kelleher, J.D.: Deep learning. MIT press (2019)
8. Li, B., Lima, D.: Facial expression recognition via resnet-50. International Journal of Cognitive Computing in Engineering 2, 57–64 (2021)
9. Luo, W., Li, J., Yang, J., Xu, W., Zhang, J.: Convolutional sparse autoencoders for image classification. IEEE transactions on neural networks and learning systems 29(7), 3289–3294 (2017)
10. Rujirakul, K., So-In, C.: Histogram equalized deep pca with elm classification for expressive face recognition. In: 2018 international workshop on advanced image technology (IWAIT). pp. 1–4. IEEE (2018)
11. Teoh, K., Ismail, R., Naziri, S., Hussin, R., Isa, M., Basir, M.: Face recognition and identification using deep learning approach. In: Journal of Physics: Conference Series. vol. 1755, p. 012006. IOP Publishing (2021)
12. Tolba, A., El-Baz, A., El-Harby, A.: Face recognition: A literature review. International Journal of Signal Processing 2(2), 88–103 (2006)
13. Wang, H., Li, S.Z., Wang, Y.: Face recognition under varying lighting conditions using self quotient image. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. pp. 819–824. IEEE (2004)
14. Wiatowski, T., Bölcskei, H.: A mathematical theory of deep convolutional neural networks for feature extraction. IEEE Transactions on Information Theory 64(3), 1845–1866 (2017)
15. Zhu, Y., Min, W., Jiang, S.: Attribute-guided feature learning for few-shot image recognition. IEEE Transactions on Multimedia 23, 1200–1209 (2020)