

# Eye Detection-Based Deep Belief Neural Networks and Speeded-Up Robust Feature Algorithm

Zahraa Tarek<sup>1</sup>, Samaa M. Shohieb<sup>1,\*</sup>, Abdelghafar M. Elhady<sup>2</sup>, El-Sayed M. El-kenawy<sup>3</sup> and Mahmoud Y. Shams<sup>4</sup>

<sup>1</sup>Faculty of Computers and Information, Mansoura University, Mansoura, 35516, Egypt

<sup>2</sup>Deanship of Scientific Research, Umm Al-Qura University, Makkah, 21955, KSA

<sup>3</sup>Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt

<sup>4</sup>Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, 33511, Egypt

\*Corresponding Author: Samaa M. Shohieb. Email: sm\_shohieb@mans.edu.eg

Received: 06 July 2022; Accepted: 22 September 2022

**Abstract:** The ability to detect and localize the human eye is critical for use in security applications and human identification and verification systems. This is because eye recognition algorithms have multiple challenges, such as multi-pose variations, ocular parts, and illumination. Moreover, the modern security applications fail to detect facial expressions from eye images. In this paper, a Speeded-Up Roust Feature (SURF) Algorithm was utilized to localize the face images of the enrolled subjects. We highlighted on eye and pupil parts to be detected based on SURF, Hough Circle Transform (HCT), and Local Binary Pattern (LBP). Afterward, Deep Belief Neural Networks (DBNN) were used to classify the input features results from the SURF algorithm. We further determined the correctly and wrongly classified subjects using a confusion matrix with two class labels to classify people whose eye images are correctly detected. We apply Stochastic Gradient Descent (SGD) optimizer to address the overfitting problem, and the hyper-parameters are fine-tuned based on the applied DBNN. The accuracy of the proposed system is determined based on SURF, LBP, and DBNN classifier achieved 95.54% for the ORL dataset, 94.07% for the BioID, and 96.20% for the CASIA-V5 dataset. The proposed approach is more reliable and more advanced when compared with state-of-the-art algorithms.

**Keywords:** Eye localization; classification; DBNN; detection; feature extraction; LBP; feature extraction; SURF

## 1 Introduction

One of the most current difficulties in biometric identification systems is eye localization and recognition. This is because of its distinctive characteristics that need more precise and accurate sensors. Furthermore, efficient eye localization algorithms require face alignment, facial expression analysis, gaze estimation, and behavior analysis [1]. Therefore, eye localization has been the subject of research, despite



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the difficulties presented by factors such as variable illumination and facial expressions, as well as limited resolution and obstruction (e.g., closing eyes and wearing glasses).

Most existing works highlighted the eye pupil to be localized and detected to isolate the black pupil from the whole eye image [2]. However, extreme head movements cause various faults in the detected eye and face images, including the eye disappearing and ocular parts of the face image [3]. Furthermore, the random movement of eyes and the highlighted illumination can affect the accuracy of the detected eye image.

Currently, there are many applications that depend on the eye detection and localization, such as the application utilized to determine the driver's behavior needed to be monitored through his eye in terms of kinematic features using gaze detection algorithms. Furthermore, it is necessary to identify and authenticate people based only on their eyes, as in cases of wearing masks during the current COVID-19 epidemic.

Whether in the banks or any organization, eye localization is necessary to recognize the enrolled individuals accomplished quickly and easily without any privacy or security problems. Because of the problems listed above, eye localization has become more important and more demanded [4].

Deep Learning (DL) has shown extraordinary success across a wide range of application domains in recent years. Furthermore, Deep Neural Networks (DNNs) are primarily an effective tool for extracting features and classifying objects in computer vision and pattern recognition [5]. Deep belief networks (DBN), a kind of deep learning neural network, have been the subject of a significant amount of research and have been applied to industrial modeling processes and extracting deep features [6].

With feature-extraction capabilities and robust hierarchical learning, deep neural networks built on DBNs may be concerned as the forecasting approach for achieving high predictive accuracy by employing effective characteristics of practical-complex data [7].

This paper proposes the Speed Up Robust Features (SURF) Algorithm to reduce the dimensional feature vectors and the elapsed time required for matching. The SURF algorithm is responsible for extracting face images using the Hessian matrix and Haar wavelet transform based on the response results from the window descriptor [8].

Even though SURF's discrimination rate is high, it consumes elapsed time lower than other methods that have been used in the past. Alternatively, there is no need for a huge number of bits in the feature vectors, so the comparison time is shorter and quicker [9]. This proves the ability to improve the accuracy of the total measurement.

The local binary pattern (LBP) is an effective non-parametric factor that may be applied to describe local image characteristics. The center pixel value determined it compared with the neighborhood pixel values, which is made by matching the central pixel's grey value with the grey values of the eight surrounding pixels [10].

The complex eye appearance is caused by changes in the shape, size, color, and location of the iris. Moreover, eyelids and eyelashes produce external noises referring to various factors, including but not limited to hair, shadow, eyeglasses, and low-quality images produced by sensors [11].

In this paper, we use the SURF algorithm to localize facial components quickly, particularly the eye and pupil of the whole face image. The anterior box of facial components is recognized by eye positioning based on Hough Circle Transform (HCT) and Local Binary Pattern (LBP). After that, the classification is carried out using DBNN based on Restricted Boltzmann Machine.

This work's novelty is utilizing DBNN and SURF with LBP to detect different parts of face images with reliable and promising results. Moreover, the ability of the proposed system to learn and extract the features from different subjects in the presence of low-quality images and resolution. In this work, we further

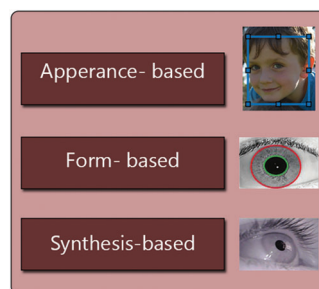
determine the detection rate results, Accuracy, Precision, Recall, Sensitivity, F1-score, and the elapsed time of the detected eye images. The main contributions of the proposed study are listed as follows:

- Face detection and eye localization using LBP and SURF algorithm,
- Classification of the detected eye images using DBNN,
- Applying Stochastic Gradient Descent optimizer and fine-tuning the hyper-parameter of DBNN,
- Applying several datasets, namely BioID, ORL, and CASIA-V5,
- Comparison between the proposed model and other recent approaches.

The remainder of the paper is organized as in the following sections: Section 2 describes the relevant work of the most modern eye detection methods. In Section 3, we provide the proposed DBNN architecture using SURF and HCT for localization of the enrolled eye images and then classify the successfully and incorrectly detected eye images. Section 4 investigates the proposed algorithm's results and discussion. In Section 5, the conclusion and perspectives are illustrated.

## 2 Related Work

Facial alignment, gaze estimation, driver tiredness detection, and many other applications rely on precisely assessing eye-image extracted features [12]. As shown in Fig. 1, the most recent approaches in eye localization have concentrated on three distinct categories. The first category is called appearance-based, by which the detection of facial images is performed. The second category is form-based to identify the different parts of eye images, such as the pupil, eyelid, and eyelashes using segmentation and localization approaches. The extracted features of eye images are stored as templates; these templates are then applied to the matching process by comparing the enrolled template with the restored templates in the database. The last category is called synthesis-based, which is generated by merging shape-based and appearance-based models. However, synthesis-based strategies use the most relevant parts of one model and ignore the redundant parts of the other options [13].



**Figure 1:** The most recent localization techniques

Ahmad et al. [14] suggested Faster Recurrent Convolutional Neural Networks (FRCNN), AlexNet, and a Rectangular-Intensity-Gradient (RIG) technique as a unified method for face recognition, eye recognition, detection of eye openness, and eye center localization. Both eyes were identified using Faster RCNN, then, AlexNet assisted in determining the eye's status (closed or open). Thirdly, the eye's center was recognized by the RIG technique.

Padmapriya et al. [15] presented an automated eye localization approach based on Infrared (IR) thermal pictures (not visual images) and the YOLO-v2 object detector. For test images, the mean average accuracy was 97%, while the mean Intersection over Union (IoU) was 90%.

In [2], they used a Generative Adversarial Network (GAN) to solve the problem of inadequate representation in multi-view face images. They used the ripple-iris filter to achieve greater precision in gradient localization. The findings showed that their approach achieved superior accuracy and more resilience in eye localization when using four different applied datasets.

In [16] proposed, a hybrid regression technique known as the Supervised Descent Method (SDM) to localize low-resolution eye images using two public datasets. The experiment results demonstrated that the approach significantly increased performance in terms of precision and accuracy as well as a lower level of complexity.

In [17] provided a standard database to localize iris images using training and testing phases based on ResNet-50, MobileNet-V2, VGG-19, and VGG-16 Transfer Learning (TL) classifiers. These TL classifiers use an iris bench-mark dataset, and ImageNet weights were utilized to initialize the models in the training phase.

Table 1 lists some of the intelligent techniques used for eye localization [18–24]. The authors used a Weighted Binarization Cascaded Convolutional Network (WBCCN) to localize the eye image with a minimum detection error rate of 0.66.

**Table 1:** A comparative study of intelligent techniques for eye localization

Author	Technique	Purpose	Dataset	Results (%)
Liu et al. [18]	Weight Binarization Cascade Convolution Neural Network (WBCCN)	Eye localization	LFW, the web, BioID, and LFPW	Average detection errors = 0.66 (LFW), = 0.71
Abbasi et al. [19]	Particle filter method and genetic algorithm	Pupil detection	CASIA-IrisV3 and LPW	Average detection rate = 79.89%
Choi et al. [20]	GAN network	Eye pupil localization	BioID and Gi4E	Accuracy <sub>(BioID)</sub> = 93.3 Accuracy <sub>(GI4E)</sub> = 99.6
Ahmed et al. [12]	CNN, Explicit Shape Regression (ESR)	Iris Center (IC) and Eye Corner (EC) localization	BioID and Gi4E	Accuracy = 97.44
Ahmed [21]	Rough-to-fine eye ROI approach and image gradient-based algorithm	Detecting the center of the eye (COE)	GI4E, BioID, and face video datasets	Accuracy <sub>(BioID)</sub> = 91.68 Accuracy <sub>(GI4E)</sub> = 96.70
Pan et al. [22]	Morphology operations, In the annular region, hough transform	Iris localization	Casia-V3.0-Interval, Lamp, and Twins	Accuracy = 95.00
Winston et al. [23]	CNN with SVM	Iris recognition	IIT Delhi iris	Accuracy = 97.80
Xue et al. [24]	Random forest	Eye center localization	BioID	Accuracy = 92.88

While in [19], a particle filter method-based Genetic Algorithm (GA) is presented to detect the black pupil from an eye image using CASIA-V3 with an average accuracy of 79.89%. Choi et al. [20] proposed an eye pupil localization approach using the BioID dataset, and they utilized a GAN network to generate the eye images with an average accuracy of 93.30%.

Furthermore, Ahmed [21] presented an algorithm to detect the Center of the Eye (CoE) using the Eye Region of Interest (ERoI) based on BioID, and the accuracy achieved was 91.68%.

The morphology operations of the annular region of the eye image based on the Hough transform were presented by Pan et al. [22] to localize the iris image using the CASIA-V3 dataset, and the resulting accuracy was 95.00%. Moreover, Winston et al. [23] utilized Convolutional Neural Networks (CNN) with a Support Vector Machine (SVM) for iris recognition based on the IIT Delhi iris dataset with 97.80% accuracy. In [24] presented an Eye Center Localization (ECL) approach applied to the BioID dataset with an accuracy of 92.88%.

This work employed a combination of LBP and SURF methods to recognize and localize eye regions in facial images. Consequently, we determine the Region of Interest (RoI) of the ocular area and save each resulting template in a database. The stored templates results from the extracted SURF features are then applied to the classification stage using DBNN with a fine-tuning hyperparameter value and SGD optimization. Through the training process of a DBNN, we analyzed the enrolled eye images and determined the degree of similarity and matching score values for each iris template.

### 3 Proposed Method

The essential objective of this study is to recognize eye images from whole face images and consider gaze estimation, multi-pose variations, lighting, and images with poor quality. While tracking previous work, a little effort was utilized to tackle the mentioned challenges, especially low-resolution images and multi-pose face recognition. In this work, the proposed algorithm consists of the following steps:

- Detection of the face and extraction of the ERoI.
- Localization of iris and pupil using SURF and LBP.
- Extracted features normalization.
- Template generation for feature vectors.
- Saving the generated template in the database.
- Classification of the enrolled templates using DBNN.
- Training and testing of the previously stored templates.

These phases rely on data preprocessing, feature extraction, encrypting, and matching, as listed below.

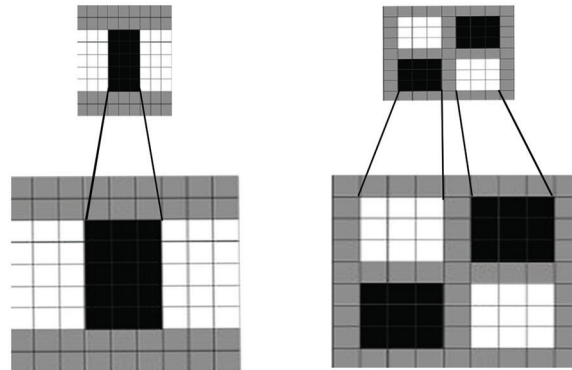
#### 3.1 Preprocessing and Extraction of Features

The obtained facial images include some significant information, such as the top location of the eye in the image, which includes the eyelids, pupil, and sclera. Nevertheless, the primary function here is to identify the eye area itself. As a result, we utilized SURF and LBP methods to find and detect the eye in the entire face. After that, we converted the image into the proper format to extract the necessary characteristics. Here are the most important preprocessing steps, whose order is set by LBP and SURF:

##### 3.1.1 Speeded Up Robust Features (SURF)

The eye's image is derived from the face by first locating the pupil's circle, then detecting the region of interest (ROI) from the face image. In addition, this model does not consider the junction of the iris with the eyelids and the reflections and eyelashes. The SURF algorithm is widely regarded as among the most powerful and trustworthy face and eye identification algorithms available today. Based on our available

information, the SURF method is three times quicker than the Scale-Invariant Feature Transform (SIFT), and it also has good recall and accuracy performance. In addition, SURF is far more helpful in overcoming the obstacles of face identification faced by multi-poses and rotation [25]. Fig. 2 describes the structure of SURF.



**Figure 2:** Speeded up robust features structure

The SURF detection method is based on the Hessian interest point (HIP) in such a way that the HIP matrix controls the localization process, as shown in Eq. (1).

$$HIP = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{yx} & L_{yy} \end{bmatrix} \quad (1)$$

where  $L_{xx}$  represents the eye images in terms of its Gaussian Laplacian. The HIP is computed using the convolution of a Gaussian with the derivative of the second order of the restored eye templates. The approximate values of HIP parameters  $L_{xx}$ ,  $L_{xy}$ ,  $L_{yx}$ ,  $L_{yy}$  are calculated by  $D_{xx}$ ,  $D_{yy}$ ,  $D_{yy}$  and  $D_{xy}$  as given in Eq. (2) such that:

$$Det(HIP_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \quad (2)$$

where  $w$  represents the rectangle region's weights. Utilizing SURF and spatial filter analysis for low-quality images and noise removal is required before the matching procedure. Du et al. [26] provided a way to detect faces based on the SURF algorithm having 64 and 128 SIFT features. The findings indicated that the SURF technique is faster than the SIFT method. Therefore, SURF can accelerate the detection process by employing quick indexing via the Laplacian parameters for the most relevant ocular features of eye images supplied by the Trace  $T$  of the Hessian matrix, as in Eq. (3).

$$T = L_{xx} + L_{yy} \quad (3)$$

### 3.1.2 Local Binary Pattern (LBP)

LBP is an effective technique for extracting the features from grayscale images. The LBP was developed as a pattern assessment for invariant grayscale images and a scale connected to the contrast of the local image. LBP is favoured for usage in fields requiring categorization of textures and rapid feature extraction. Image retrieval and biological image analysis are one of the important applications using the LBP texture operator to extract the most relevant details required for diagnosis and prognosis.

LBP is introduced in this work to retrieve the features of the eye images from the normalized ERoI. The DBNN classifier receives the feature vectors as an output from the LBP algorithm, each of which has  $n$

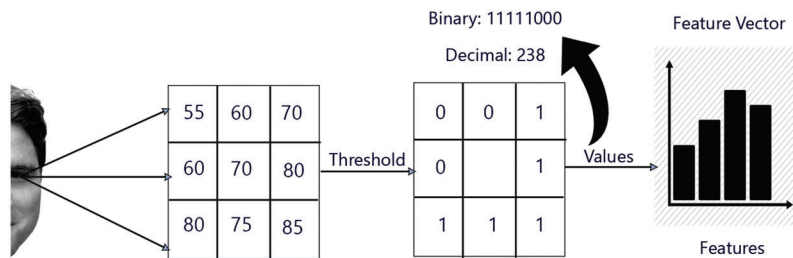
dimensions and then the LBP codes are generated by employing the sampling points P with a circle radius R, given an intensity image I, as shown in Eqs. (4) and (5) [27].

$$LBP = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (4)$$

$$s(g_p - g_{pc}) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (5)$$

where  $g_c$  denotes the center pixel's gray value and  $g_p$  denotes the  $P^{\text{th}}$  neighbor's gray value.

After the eye area has been localized, it is converted into a template with a predetermined length, compared to the prestored eye images. But there is still the problem of contrast in iris templates such as the variance problem of iris histogram. There are several reasons for the variance, including the change of image distance, camera rotations, and head tilt [28]. Fig. 3 illustrates the LBP process.



**Figure 3:** Local binary patterns process

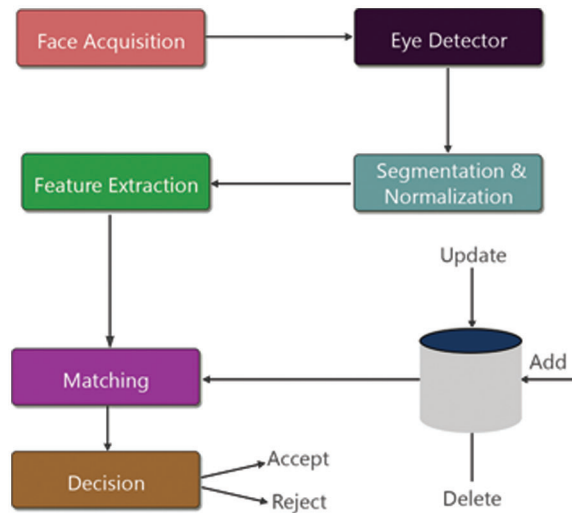
In this work, it was necessary to use the Hough circle transform (HCT) in order to get a polar coordinate from the Cartesian coordinate. Additionally, z-score and min-max normalization are used to improve the eye templates. The encoding and feature extraction stages are utilized as shown in Fig. 3 by which the most significant features of the eye image are extracted using LBP.

### 3.2 Matching

The matching process usually compares the template entered into the system with the templates previously stored in the database. In this paper, the face acquisition process begins with capturing and acquiring images of the face. Next, we apply SURF and LBP to localize the eye images' different parts, including the eye, nose, and mouth parts, as shown in Fig. 4. Segmentation of eye images is used to focus on the EROI and to discard the redundancy. Hence, iris normalization based on z-Score normalization is performed to obtain the template. Extracted templates require feature extraction to get feature vectors. This feature vector is a template stored in the database. When a person records their eye image, they will add or update their eye templates. The matching stage then determines the similarity between the input iris template and the template stored in the database which is called identification process to get the decision to accept or reject the registered person depending on the matching score values [29].

### 3.3 Classification Based-DBNN

Deep Belief Neural Networks (DBNs) are made up of Sigmoid Belief Networks (SBNs) and Restricted Boltzmann Machines (RBMs), which are both unsupervised learning networks. The most significant benefit of DBNs is their capacity to acquire knowledge about learning extracted characteristics step by step, using learning techniques that enable higher-level characteristics to be learned from lower-level ones [30]. We employed DBN for classification in this study, which involves layer creation and connection to guarantee that the graphical model consists of numerous layers for the latent variables.



**Figure 4:** The matching process of the eye identification system

This study's proposed DBN architecture consists of two Restricted Boltzmann machines (RBMs) layers and one Multi-Layer Perceptron (MLP). Furthermore, RBM inference is quite similar to single-layer neural network inference [31–33]. The vector representation of a neural network with a single layer is described in Eq. (6).

$$h = \sigma(w^T o + b) \quad (6)$$

where  $\sigma(x)$  represents the sigmoid function,  $w$  is the weight,  $o$  is a recognized visual characteristic, and  $b$  is a constant. Fig. 5 provides the proposed DBNN structure, which comprises two RBM and one MLP. The visible layer enrolls the observed input characteristic designated by  $O_n$  in the first RBM and passes the outcomes by modifying the weights  $w^1$  to the hidden layer. By adjusting the weights  $w^1$  to enhance the enrolled features through the RBM-2, the outcome of the first RBM fed into the second RBM as observed input values. The features are received by the MLP, which then passes via the hidden layers to produce  $m$  groups of observed data according to their Euclidean Distance (ED).

Algorithm 1 summarizes all phases throughout the classification stages and shows the stages of the proposed DBNN. In Fig. 6, all phases of the present model based on hybrid DBNN, SURF, and LBP are shown, along with the corrected detected eye images' matching method. Also, we use a matching technique to decide if whether the registered templates of the eye photos can be appropriately detected depending on a threshold values that has been set in advance, as shown in Fig. 6.

The approach that compares the pre-stored forms in the dataset with the registered templates is referred to as the one-to-one verification method. This comparison helps identify the template error values. That can ascertain the position of the saved templates and extract them in a time-efficient manner. In addition, the capability of the proposed system to detect the error values of the improperly matched eye template is dependent on a certain threshold that it is determined using trial and error methodology.



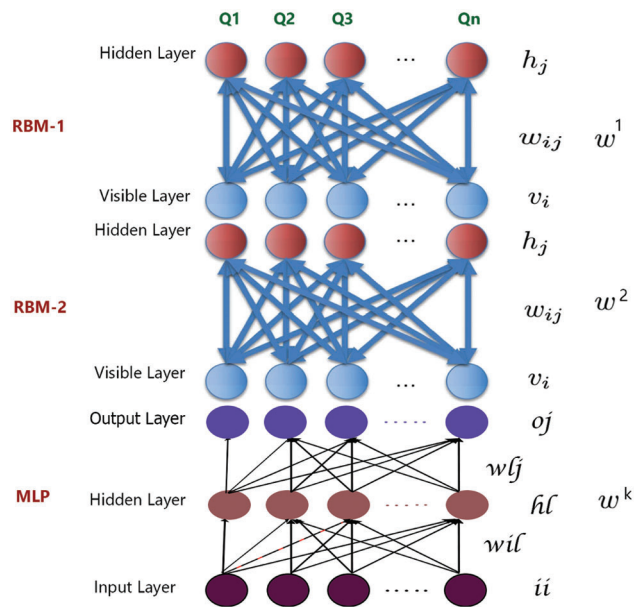


Figure 5: DBNN's proposed architecture

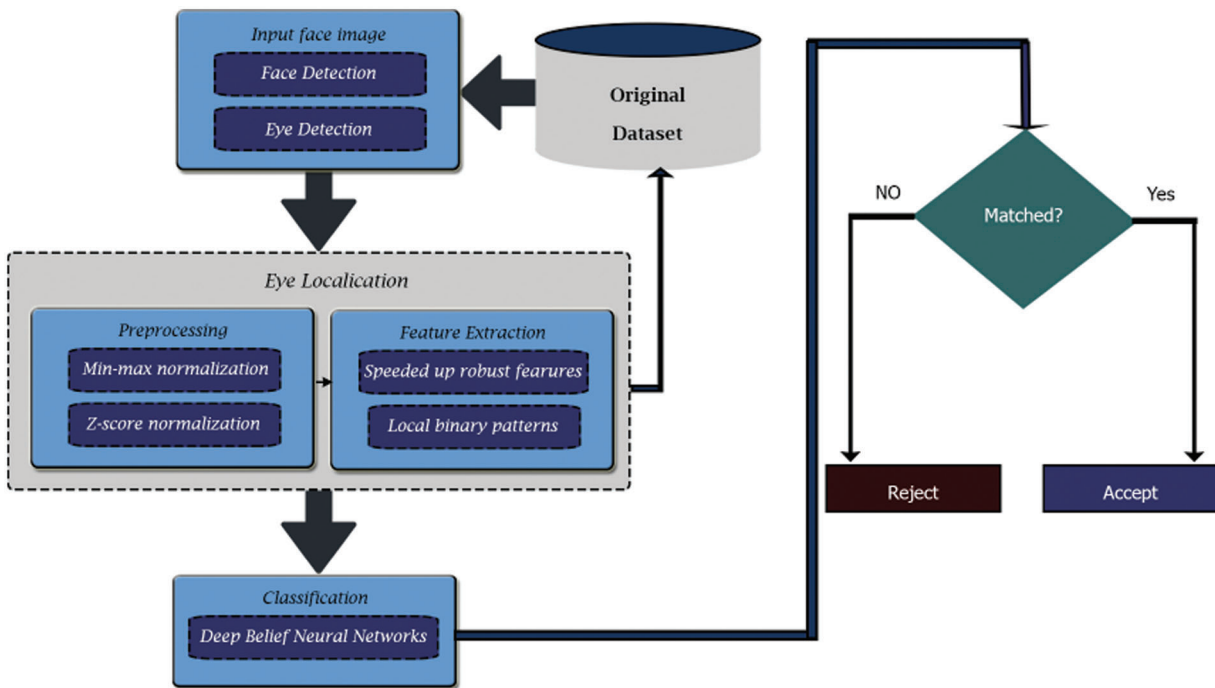


Figure 6: The proposed hybrid LBP, SURF, and DBNN algorithms' flow chart

---

**Algorithm 1:** The proposed DBNN classification stages.

---

**Input:** initializing the retrieved feature vectors from SURF and LBP. Let  $I$  represent the number of individuals, then the extracted characteristics of eye images are provided by:

$$O_x = F_x = \{F_{x1}, F_{x2} \dots F_{xi}\}_{i=1}^I.$$

**Initialization:** Tuning the weight to zero and performing a pre-training phase on the hyper-parameters of DBNN in the RBM layers.

**Procedure:**

1. Initializing the vector representation of an RBM network with a singular layer:

$$h = \sigma(w^T o + b),$$

2. Computing the gradient descent of the observed templates to determine the information contained inside Feature Vectors (FV),
3. Updating the weights and calculating the gradient, momentum, and weight decay,
4. Determining the posterior likelihood of the inference model given the hidden variable,
5. Utilizing SoftMax with the activation variable vectors,
6. Using Euclidean Distance to figure out the dissimilarity measure that compares the query  $q_i$  to the previously recorded  $o_i$  templates.

**Output:**

1. Comparing the scores of every class,
  2. A confidence-based choice (Reject/Accept).
- 

#### 4 Results and Discussion

In this part, we present our findings with three distinct facial image datasets: ORL [34], BioID [35], and the CASIA-V5 dataset [36]. The ORL database includes 40 individuals with 400 face images generated at the AT & T laboratory at the University of Cambridge. BioID comprises 1521 face images with a consistent resolution of  $384 \times 286$  pixels with an anterior perspective of participants' faces. Lastly, CASIA-V5 comprises 500 individuals with 2,500 face images in various positions. Table 2 provides a description of the collected dataset and the total number of images that were trained and evaluated, with 70 percent for training and 30 percent for testing, respectively.

**Table 2:** The characteristics of the datasets

Dataset	Dataset characteristics			
	# Images	# Trained samples	# Tested samples	Number of subjects
ORL	400	280	120	40
BioID	1521	1065	456	23
CASIA-V5	2500	1750	750	500
Total	4421	3095	1326	563

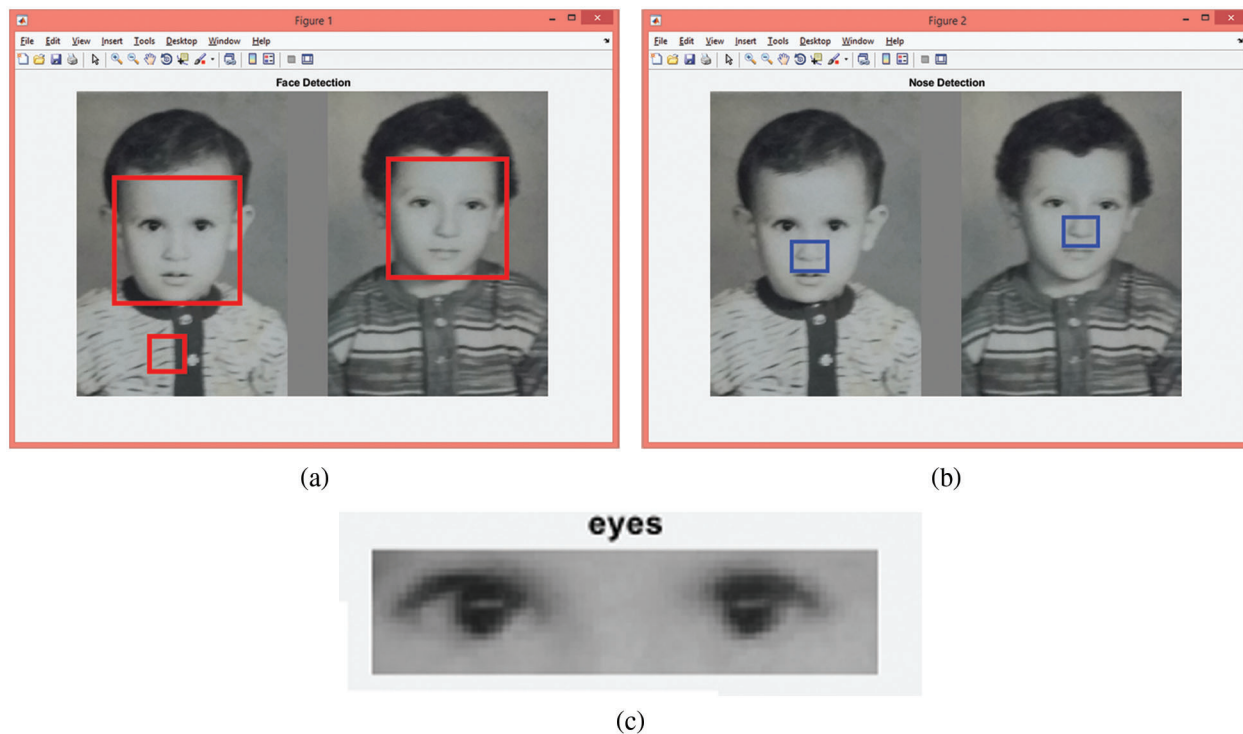
The experiment results depend on the fixed values of the hyper-parameters shown in Table 3. A feature vector of length 512 is used to perform normalization on the registered eye templates included within the

dataset. The number of iterations is 250. The initial momentum is 0.5 of what it will be in the end, which is 0.90. Moreover, the learning rate and the weighted cost are 0.01 and 0.002, respectively.

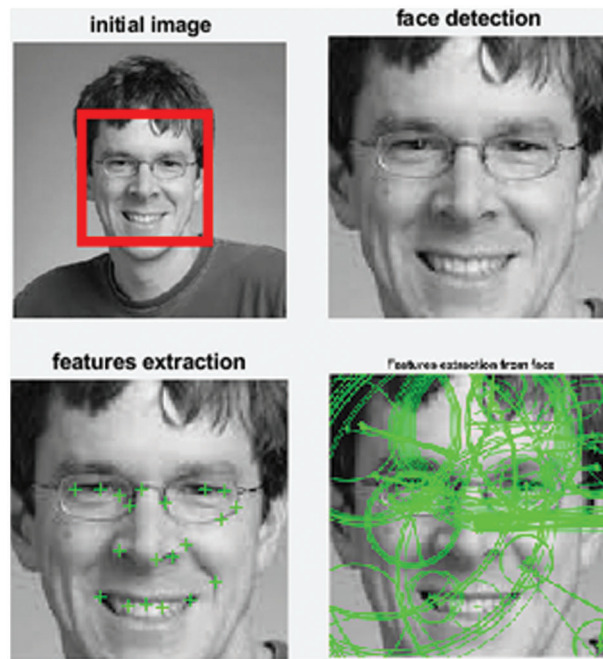
**Table 3:** DBNN’s hyper-parameters values

Hyperparameters	Values
Feature vector length	512
Number of iterations	250
Initial and Final momentum	0.5, 0.9
Learning rate	0.01
Weight cost	0.002

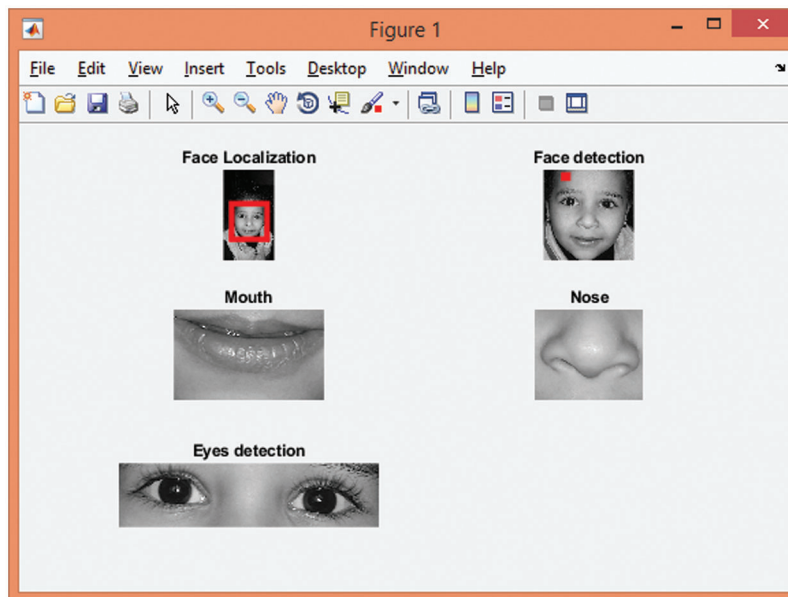
The findings of this research imply that the localization of ocular images is based not only on LBP and SURF, but also on the concatenated normalized forms of SURF and LBP. Fig. 7a, depicts the localization of facial images for two individuals simultaneously using SURF. While in Fig. 7b, the location of each subject’s nose is shown. Fig. 7c, shows the localization of the eye region using SURF, which was acquired from the image of the full face of a single individual. Fig. 8 is a composite image depicting the extracted characteristics of facial images, containing the facial localization procedure. The dots indicate the extracted features and distances of face images using the SURF algorithm. Fig. 9 is a cluster image of how SURF may extract all face characteristics simultaneously. Fig. 10 illustrates the localization of eye samples from facial image and iris identification using HCT and LBP, as well as the detection of edges using the Canny approach.



**Figure 7:** Using the SURF algorithm, position of (a) Face detection, (b) Nose detection, and (c) eye detection



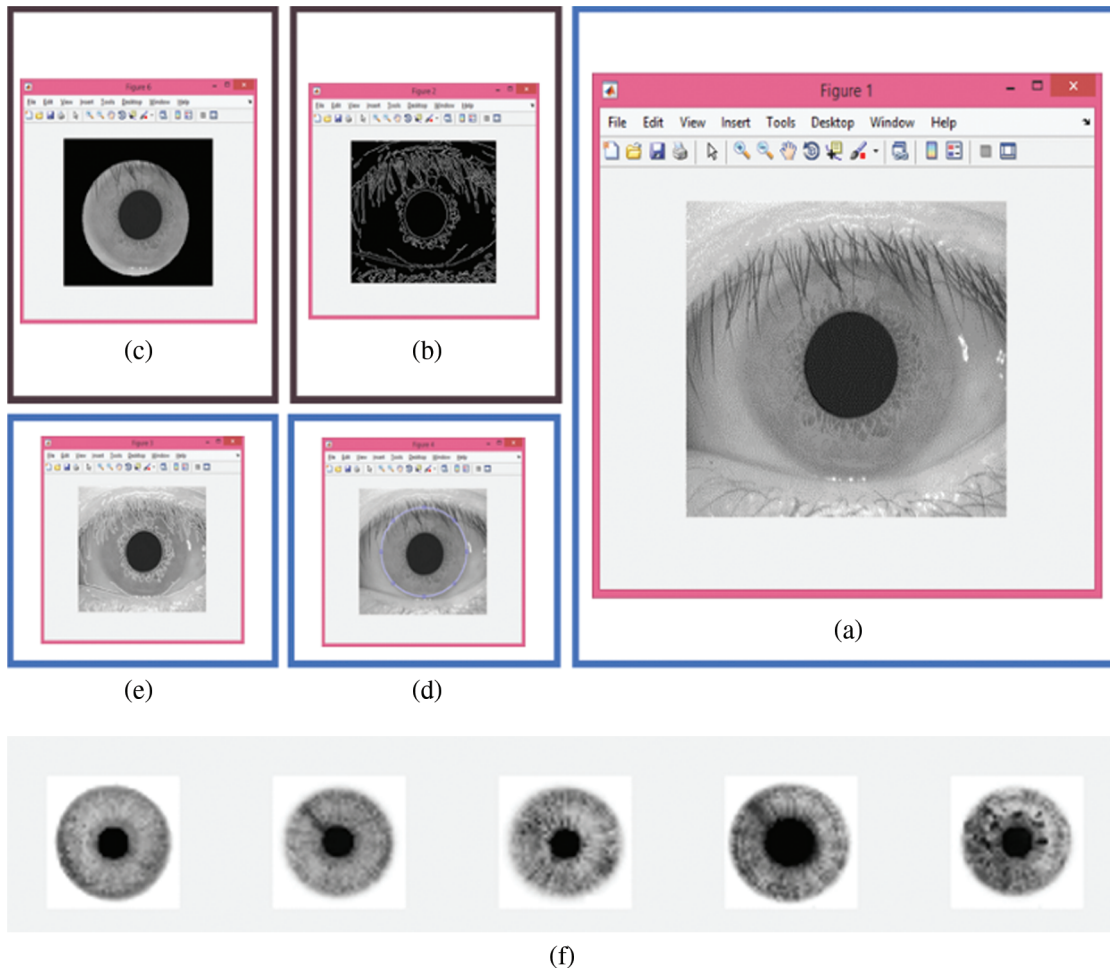
**Figure 8:** Face localization, detection, and feature extraction with the SURF method



**Figure 9:** Face localization, detection, mouth, nose, and eyes detection using poor enrolled images quality

Using Euclidean distances to establish matching scores, assessment results are obtained. In addition, the detection rate accuracy is determined based on the number of iterations to realize its final value, as seen in Fig. 11. The number of iterations affects the detection rate of outcomes, as shown in Fig. 11a, and the results from the ORL dataset yield a 95.54 percent success rate based on LBP, SURF, and DBNN. The CASIA-V5 and BioID detection rates were 96.20% and 94.07%, respectively (Figs. 11b and 11c). The curve gives better results and stability when reaching 250 iterations. While in 100 iterations, there is no stability

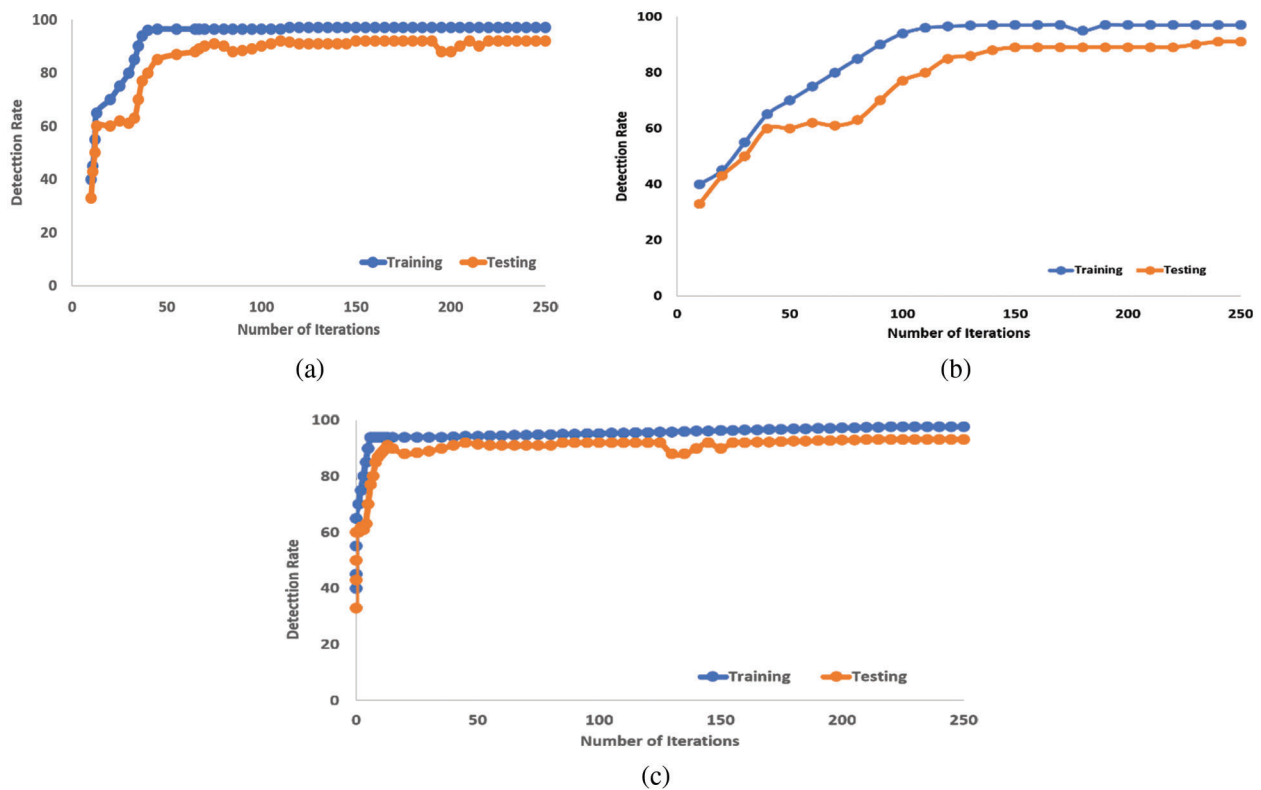
as there is a slight decrease in the results. Therefore, we utilized the SGD optimizer to tackle the over-fitting and to stabilize the accuracy at 250 iterations.



**Figure 10:** Eye image localization and segmentation using HCT and LBP. (a) Initial eye sample, (b) Edge detection by Canny, (c) Binary mask for segmentation, (d) Iris localization, (e) Enrolled eye image, and (f) Enrolled eye templates

Detection rates using SURF, LBP, SURF + LBP, and DBNN are all examined in [Table 4](#) for the three different applied datasets. We found that the hybrid methods of LBP, SURF, and DBNN yield better precision than the use of any of these systems alone. This is due to the proposed system's ability to make use of the strengths of each algorithm while minimizing the drawbacks of the others. According to the results provided in [Figs. 12](#) and [13](#), we compare the suggested strategy to the Convolutional Neural Networks (CNN), Support Vector Machine (SVM), and Radial Basis Function (RBF) in terms of the detection rate ([Fig. 12](#)) and the elapsed time ([Fig. 13](#)). The same hyperparameter values indicated in [Table 3](#) were used to compare the datasets. SVMs, RBFs, and CNNs were all compared to show that the proposed system has a superior detection rate accuracy than the others. On the contrary, the proposed system has a longer elapsed time than the alternatives. As such, the training procedure based on DBNN needed additional time since it had to duplicate some of the startup phases described in an [Algorithm 1](#) to update weights

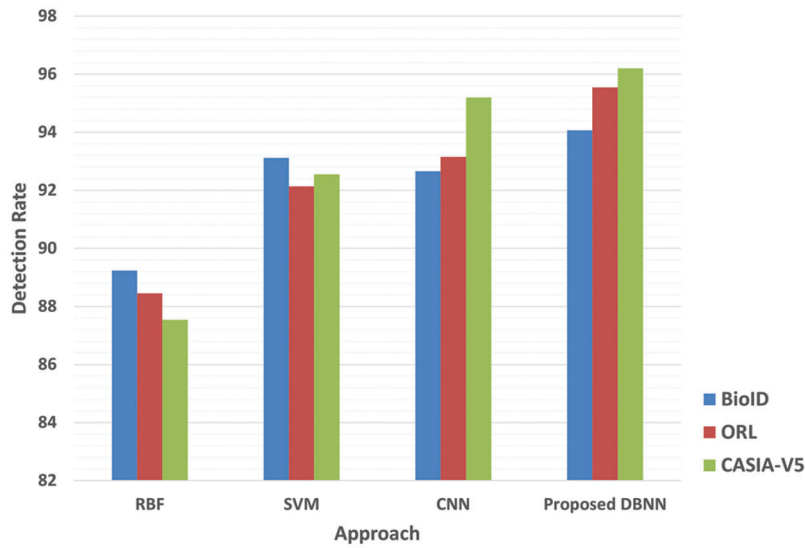
and pre-train the models. We employed a stochastic gradient decent operator to combat the issue of overfitting, but the elapsed time remained greater. According to our best understanding, CNNs may not be able to reliably locate and identify objects in low-resolution photos, multiple-pose images, or in low-light conditions since they lack the ability to accurately recognize features. As a result, we have developed a hybrid detection and localization approach that combines SURF and LBP as a constant template that is utilized for classification based on DBNN. This method allows us to pinpoint the location of an eye more accurately. To boost the results obtained, we used the confusion matrix to determine the accuracy, recall, precision, and the harmonic mean of precision and sensitivity, which is called the F1-score as shown in Eqs. (7)–(10) [37,38].



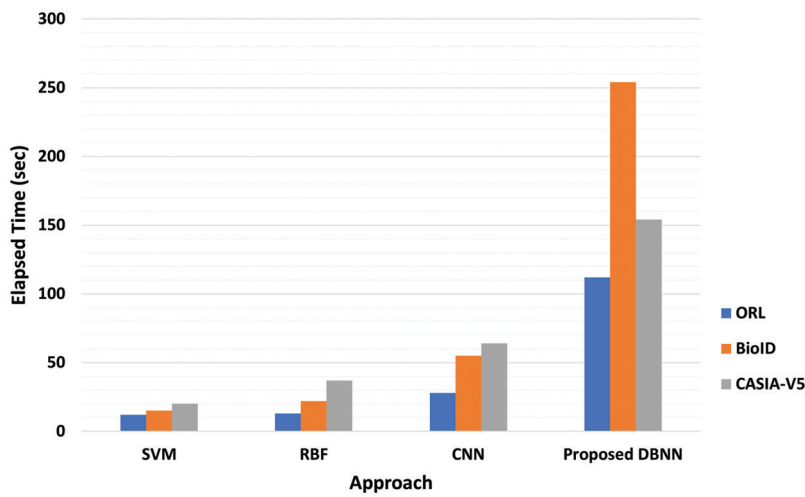
**Figure 11:** Proposed DBNN’s accuracy dependent on LBP and SURF. (a) ORL, (b) BioID, and (c) CASIA-V5 dataset

**Table 4:** The detection rates for the suggested SURF + DBNN, LBP + DBNN, and LBP + SURF + DBNN combinations

Approach	Detection rate (%)		
	ORL	BioID	CASIA-V5
SURF + DBNN	92.02	93.01	94.54
LBP + DBNN	91.00	92.00	93.20
LBP + SURF + DBNN	95.54	94.07	96.20



**Figure 12:** The comparative examination of CNN, RBF, SVM, and proposed DBNN



**Figure 13:** The comparative analysis of the CNN, RBF, SVM, and proposed DBNN elapsed time

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

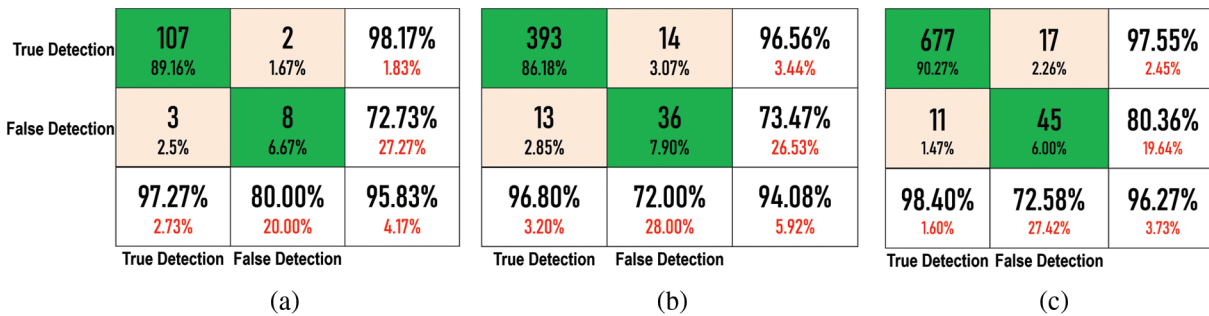
$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{10}$$

where TP is the true positive rate, the true detection rate in both actual and predicted values, TN is the correctly nondetected eye images or the true negative rate, FP, and FN are the wrongly detected rate or

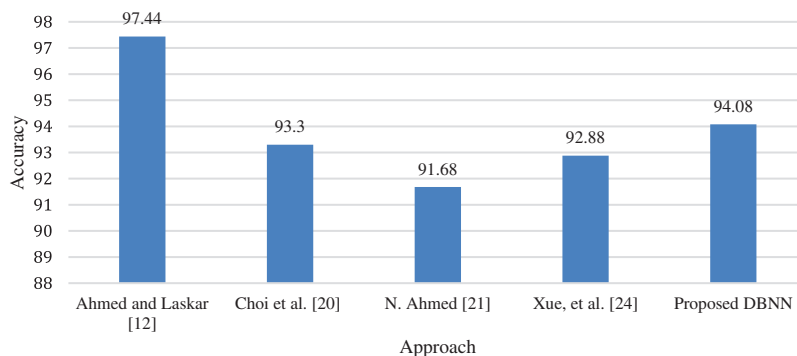
the false positive and false negative rates, respectively. Here, the experimental results are based on the correctly and wrongly detected eye images from the whole face image for the applied ORL, BioID, and CASIA-V5 datasets. Fig. 14 investigated the confusion matrices of results obtained from the applied three datasets using the LBP and SURF with DBNN. Table 5 shows the accuracy, recall, precision, and F1-score of the obtained results in the testing phase. The comparison between the proposed DBNN and [12,20,21,24] using the applied BioID dataset is shown in Fig. 15. Although in [12], the detection accuracy is higher than the proposed DBNN, using the SGD can achieve more accurate results. Furthermore, the proposed DBNN can be utilized using different standard datasets and any image format, even low-quality enrolled images.



**Figure 14:** The confusion matrices of the proposed LBP + SURF + DBNN in testing phase for the applied (a) ORL, (b) BioID, and (c) CASIA-V5 dataset

**Table 5:** The accuracy, recall, precision, and F1-score of the proposed method for the applied datasets in the testing phase

Dataset used	Accuracy	Recall	Precision	F1-score
ORL	95.83	98.17	97.27	97.71
BioID	94.08	96.56	96.80	96.68
CASIA-V5	96.27	97.55	98.40	97.97



**Figure 15:** The comparative study of the proposed DBNN with the recent approaches [12,20,21,24] using the BioID dataset



## 5 Conclusion and Perspectives

Many applications, including privacy, health, legal, and authentication, continue to rely on eye detection as a primary difficulty. This study's key obstacles and inspirations include the wide variety of poses and illumination conditions and the poor-quality of the facial images. According to the findings of this work, the proposed localization system using LBP and SURF based on DBNN achieved a higher detection rate accuracy than standard SVM, RBF, and the CNN classifier. After SURF has been used to identify the face, HCT and LBP used to locate the pupil area of the eye in the face image. Furthermore, the DBNN showed promise in its ability to identify the detected images due to the combination of the LBP and SURF algorithms, which produced promising results. An observed drawback of this work is that it takes longer elapsed time to train the images within the DBNN than other methods. Therefore, we can improve the results by using one or more quantization algorithms, e.g., Vector Quantization (VQ) and dimensionality reduction algorithms such as Principal Component Analysis (PCA). This leads to an improvement in the elapsed time due to the size minimization of the template feature vector; hence, it is possible to obtain a reduction in the elapsed time.

**Acknowledgement:** The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (22UQU4331164DSR02).

**Funding Statement:** The author(s) received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] R. S. Hessels, G. A. Holleman, A. Kingstone, I. T. C. Hooge and C. Kemner, "Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors," *Cognition*, vol. 184, pp. 28–43, 2019.
- [2] W. -Y. Hsu and C. -J. Chung, "A novel eye center localization method for multi-view faces," *Pattern Recognition*, vol. 119, no. 108078, pp. 1–13, 2021.
- [3] W. -Y. Hsu and C. -J. Chung, "A novel eye center localization method for head poses with large rotations," *IEEE Transaction Image Processing*, vol. 30, pp. 1369–1381, 2020.
- [4] B. Huang, R. Chen, Q. Zhou and W. Xu, "Eye landmarks detection via weakly supervised learning," *Pattern Recognition*, vol. 98, no. 107076, pp. 1–11, 2020.
- [5] J. Yu and G. Liu, "Knowledge transfer-based sparse deep belief network," *IEEE Transactions on Cybernetics*, vol. 52, no. 8, pp. 1–12, 2022.
- [6] G. Wang, J. Qiao, J. Bi, Q. -S. Jia and M. Zhou, "An adaptive deep belief network with sparse restricted boltzmann machines," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 4217–4228, 2019.
- [7] G. Wang, Q. -S. Jia, M. Zhou, J. Bi, J. Qiao *et al.*, "Artificial neural networks for water quality soft-sensing in wastewater treatment: A review," *Artificial Intelligence Review*, vol. 50, pp. 565–587, 2022.
- [8] M. Bansal, M. Kumar and M. Kumar, "2D object recognition techniques: State-of-the-art work," *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 1147–1161, 2021.
- [9] M. A. Taha and H. M. Ahmed, "Speeded up robust features descriptor for iris recognition systems," *Journal of University of Babylon for Pure and Applied Sciences*, vol. 29, no. 2, pp. 244–257, 2021.
- [10] A. S. Mubarak, S. Serte, F. Al-Turjman, Z. S. Ameen and M. Ozsoz, "Local binary pattern and deep learning feature extraction fusion for COVID-19 detection on computed tomography images," *Expert Systems*, vol. 39, no. 3, pp. 1–13, 2022.
- [11] M. Ahmed and R. H. Laskar, "Eye center localization in a facial image based on geometric shapes of iris and eyelid under natural variability," *Image and Vision Computing*, vol. 88, pp. 52–66, 2019.

- [12] M. Ahmed and R. H. Laskar, "Evaluation of accurate iris center and eye corner localization method in a facial image for gaze estimation," *Multimedia Systems*, vol. 27, no. 3, pp. 429–448, 2021.
- [13] M. Y. Shams, A. E. Hassanien and M. Tang, "Deep belief neural networks for eye localization based speeded up robust features and local binary pattern," in *Proc. of LISS*, China, Beijing Jiaotong University, pp. 415–430, 2022.
- [14] N. Ahmad, K. S. Yadav, M. Ahmed, R. H. Laskar and A. Hossain, "An integrated approach for eye centre localization using deep networks and rectangular-intensity-gradient technique," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 1–15, 2022.
- [15] N. Padmapriya, N. Venkateswaran, R. Ravikumar and R. Chelliah, "Localization of eye region in infrared thermal images using deep neural network," in *Proc. of Sixth Int. Conf. on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, India, pp. 446–450, 2021.
- [16] Y. Xia, J. Lou, J. Dong, L. Qi, G. Li *et al.*, "Hybrid regression and isophote curvature for accurate eye center localization," *Multimedia Tools and Applications*, vol. 79, no. 1, pp. 805–824, 2020.
- [17] M. Adnan, M. Sardaraz, M. Tahir, M. N. Dar, M. Alduailij *et al.*, "A robust framework for real-time iris landmarks detection using deep learning," *Applied Science*, vol. 12, no. 11, pp. 5700, 2022.
- [18] Z. -T. Liu, S. -H. Li, M. Wu, W. -H. Cao, M. Hao *et al.*, "Eye localization based on weight binarization cascade convolution neural network," *Neurocomputing*, vol. 378, pp. 45–53, 2020.
- [19] M. Abbasi and M. R. Khosravi, "A robust and accurate particle filter-based pupil detection method for big datasets of eye video," *Journal of Grid Computing*, vol. 18, no. 2, pp. 305–325, 2020.
- [20] J. H. Choi, K. I. Lee and B. C. Song, "Eye pupil localization algorithm using convolutional neural networks," *Multimedia Tools and Applications*, vol. 79, no. 43, pp. 32563–32574, 2020.
- [21] N. Y. Ahmed, "Real-time accurate eye center localization for low-resolution grayscale images," *Journal of Real-Time Image Processing*, vol. 18, no. 1, pp. 193–220, 2021.
- [22] M. Pan and Q. Xiong, "An improved iris localization method," *The International Arab Journal of Information Technology*, vol. 19, no. 2, pp. 173–185, 2022.
- [23] J. J. Winston, D. J. Hemanth, A. Angelopoulou and E. Kapetanios, "Hybrid deep convolutional neural models for iris image recognition," *Multimedia Tools and Applications*, vol. 81, no. 7, pp. 9481–9503, 2022.
- [24] P. Xue, C. Wang, W. Huang and G. Zhou, "ORB features and isophotes curvature information for eye center accurate localization," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 1, no. 2256005, pp. 1–6, 2022.
- [25] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [26] G. Du, F. Su and A. Cai, "Face recognition using SURF features," in *Proc. MIPPR 2009: Pattern Recognition and Computer Vision*, Yichang, China, vol. 7496, pp. 1–7, 2009.
- [27] Z. Guo, L. Zhang and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognition*, vol. 43, no. 3, pp. 706–719, 2010.
- [28] L. Masek, "Recognition of human iris patterns for biometric identification," Dissertation, Master's Thesis, University of Western, Australia, 2003.
- [29] J. Daugman, "How iris recognition works," in *Proc. the Essential Guide to Image Processing*, England, Elsevier, University of Cambridge, pp. 715–739, 2009.
- [30] G. E. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, pp. 1–2, 2009.
- [31] M. Tanaka and M. Okutomi, "A novel inference of a restricted boltzmann machine," in *Proc. 2014 22nd Int. Conf. on Pattern Recognition*, Stockholm, Sweden, pp. 1526–1531, 2014.
- [32] S. Tran, "Representation decomposition for knowledge extraction and sharing using restricted boltzmann machines," Doctoral Dissertation, City University, London, England, 2016.
- [33] M. S. Alsawadi and M. Rio, "Skeleton split strategies for spatial temporal graph convolution networks," *Computers, Materials & Continua*, vol. 71, no. 3, pp. 4643–4658, 2022.
- [34] ORL face dataset, *AT&T Laboratories Cambridge*, UK: ORL Dataset, 2001. [Online]. Available: <https://cam-orl.co.uk/facedatabase.html>.

- [35] O. Jesorsky, J. Kirchberg and R. W. Frischholz, *Robust Face Detection Using the Hausdorff Distance*, BioID AG, Berlin, Germany: BioID Face Dataset, 2001. [Online]. Available: <https://ftp.uni-erlangen.de/pub/facedb/readme>.
- [36] CASIA-V5 face dataset, *National Laboratory of Pattern Recognition (NLPR)*, Institute of Automation, Chinese Academy of Sciences (CASIA), China: CASIA-V5 Face Dataset, 2010. [Online]. Available: <http://biometrics.idealtest.org>.
- [37] S. Sarhan, A. Nasr and M. Y. Shams, “Multipose face recognition-based combined adaptive deep learning vector quantization,” *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–11, 2020.
- [38] E. Hassan, M. Y. Shams, N. A. Hikal and S. Elmougy, “A novel convolutional neural network model for malaria cell images classification,” *Computers, Materials & Continua*, vol. 72, no. 3, pp. 5889–5907, 2022.