

Research Article

Deep Neural Network for Accurate Age Group Prediction through Pupil Using the Optimized UNet Model

Swathi Gowroju ^{1,2}, Sandeep Kumar ³, Aarti ⁴ and Anshu Ghimire ⁵

¹Lovely Professional University, Phagwara, Punjab, India

²Department of Computer Science and Engineering, Sreyas Institute of Engineering and Technology, Hyderabad, India

³Department of Computer Science and Engineering, Koneru Lakshmaiah Educational Foundation, Vaddeswaram, Andhra Pradesh, India

⁴Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India

⁵Department of Computer Science and Engineering, Nepal Engineering College, Kathmandu, Nepal

Correspondence should be addressed to Anshu Ghimire; anshug@nec.edu.np

Received 3 August 2022; Revised 14 September 2022; Accepted 15 September 2022; Published 13 October 2022

Academic Editor: Dinesh Kumar Saini

Copyright © 2022 Swathi Gowroju et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Predicting age automatically from the image is a difficult task and shortening the challenge to be more concise is also a challenging task. Nevertheless, the existing implementations using manually designed features using a wide variety of input features using benchmark datasets are unsatisfactory as they suffer from unknown subject information. It is challenging to judge CNN's performance using such approaches. The proposed system performs the segmentation through UNet without using a dense layer to perform the segmentation and classification. The proposed system uses the skip connection to hold the loss at the max-pooling layer. Also, the morphological processing and probabilistic classification served as the proposed system's novelty. The proposed method used three benchmark datasets, MMU, CASIA, and UBIRIS, to experiment with building a training model and tested using various optimization techniques to perform an accurate segmentation. To further test and improve the quality of the proposed method, we experimented with random images. The proposed system's accuracy is 96% when experimented on random images of subjects collected purely for experimentation. Three optimizers, namely, Stochastic Gradient Descent, RMS Prop, and Adaptive Moment Optimizer, were experimented with in the proposed system to fit the system. The average accuracy we received using optimizers is 71.9, 84.3, and 96.0 for the loss value of 2.36, 2.30, and 1.82, respectively.

1. Introduction

Over the decade, most methods were built based on statistical models, various features of the face such as wrinkles, freckles, spots on the skin, etc. Every human has to face these changes in the body as it is uncontrollable and inevitable [1]. Aging patterns may differ due to their personalized patterns, such as makeup, individual lifestyle, environmental conditions, and health [2, 3]. These attributes may instead create difficulty in estimating the age of the person. With the introduction of FGNet, this problem has been alleviated slightly by using the age dataset. For facial images, we have anthropometric models which measure the distance between

facial points from the image [4]. These are called landmark points of the face. One can define up to 132 facial measurements from the face at different locations on it. Another model, the active appearance model, is statistically and programmatically defined landmarks on the face [5–7]. These models consider both facial and texture to predict the shape of the face. These techniques are well suited for age and gender predictions. Although these methods perform well and are used widely in several applications, they need manual measurements and classification, which takes time and effort [8]. In the early 2000s, as said in Demirjian-Chaillet et al.'s method, the individual's age estimation took 10 min.

In the past decade, deep learning techniques have been used in many implementations of images. They show their accuracy well in performing segmentation, classification, and feature extractions [9–11]. We experimented with this to perform pupil segmentation to predict the age through it computationally. Age estimation based on biometric traits is attracting many researchers. One of the safe-to-use biometric traits is the eye. Eye not only authenticates a person but also hides much information needed in most applications [12–15]. One of the best applications is the result of this research paper. We proposed a rare computer application using scarce research characteristics that gave a 98% satisfying result in the experimentation. We believed in pupil dilation and used this property to estimate the person's age computationally accurately. Few researchers experimented on the proposed prediction. Most of the predictions were analyzed in the literature survey and the results were based upon the medical predictions. The computer model-based prediction was becoming problematic in terms of accuracy as the input images are always noisy due to occlusion and eyelashes. Furthermore, face detection followed by Pupil needs proper training on individual elements of the face, which is rarely available in the market [16–18]. Hence the work done on the proposed system in terms of deep learning became tough. The proposed system worked on deep and traditional methods to segment the Pupil by training the eye model of the face and generating the masked images of it for the training, through which segmentation is achieved [19–23]. The novelty of the proposed system, when compared to the existing systems, in summary, is described in the below points.

- (1) The proposed UNet can accurately identify pupil boundaries even in low illumination and noise with occlusion.
- (2) To identify the pupil boundary accurately, we performed segmentation in the first stage, followed by morphological processing in the second stage, where we can reduce segmentation errors.
- (3) The mean error rate is reduced compared to the existing state-of-the-art methods as the system is well checked with various optimizers to identify the problem.
- (4) To eliminate the overfitting problem due to the high pixel values of the image, we implemented Leaky ReLu and modified Leaky ReLu to address the problem of vanishing Gradient and dying ReLu to make the training process quicker.

The literature survey is described in Section 2 to give the paper's road map. The proposed method is discussed in Section 3. To check the accuracy level of training with UNet, we trained the system using the most popular deep learning architectures such as ResNet 50 and VGGNet along with UNet, and the training details and the pretrained models were discussed in the first subsection of Section 3. The flowchart is explained in subsections of the proposed method. Section 4 discussed the results using all possible input operations of the proposed system. Subsections of

Section 4 describe the segmentation of images from benchmark datasets, input face images, and the live video. After obtaining the segmented Pupil, we predicted the age in the 4.2 section. The accuracy of the proposed system is compared to the existing systems in Section 6.

2. Literature Survey

We studied various methods of predicting age. Characteristics that are used to predict age are, using eyes, using wrinkle features, and using dental characteristics. Age prediction using Pupils was medically proposed in 2013 by Erbilek et al. using iris biometric. They took young and old age groups collected from the University of Notre Dame, which is publicly available. They collected iris data from all 50 subjects based on the scalar distance between the iris and Pupil. According to the ratio obtained from twelve feature points, they were categorized into three age groups less than 25, 25–60, and greater than 60. In 2004, Lanitis et al. proposed AAM based method which predicts age using geometric and texture features. This methodology has a drawback: it faces difficulty if the image illumination is improper. Gao et al. in 2009 used pixel intensity and Local binary patterns to predict the age. The Gabor filter for the image used the following equation:

$$\Psi_{\mu,v}(z) = \frac{x_{u,v}^2}{\sigma^2} e^k \left[e^i x_{u,v}^2 e^{-\sigma}, 2 \right], \quad (1)$$

where $x_{u,v} = x_v e^{i\phi}$ and $\phi_u = (\pi\mu/8)$. In order to define the age membership function, he used LPB, which defines the third-order derivative of μ . The author classified the person using SVM into three classes, baby, child, adult and old. The accuracy of the proposed classification was 78.9%. As an extension to work, Guo et al. used the Gabor filter defined above with a small size, using the following equation:

$$4(x, y) = e \left(-\frac{(x^2 + y^2)}{q\sigma^2} \right) \times \cos \left(\frac{q\pi}{\lambda} k \right), \quad (2)$$

where q is 2., $x \cos \theta + y \sin \theta$ and $y = -x \cos \theta + y \cos \theta$ which rotates the filter with an angle of θ at different intervals. To decrease the further dimension of the input image, the author used PCA (principal component analysis), which works well for bio-inspired features. The classification was made for 0–9, 10–19, 20–29, 30–39, 40–49, 50–59, and 60–69 age groups using their own dataset of 1002 images, with an accuracy of 79%. Agbo-Ajala et al. proposed a deep learning classifier to predict the age and gender of the subject using an unfiltered face. The author used a CNN model using landmarks detection and face alignment to predict age, and the binary classification is calculated using the following equation:

$$H_p(x) = \frac{1}{k} \sum_{i=1}^n \left[R_j \log(p(\dot{x}_j)) + (1 - x_j) \log(1 - P(\dot{x}_j)) \right], \quad (3)$$

where x is the binary class label for predicted points on the face, IMDB -WIKI and MORPH-II datasets were used in the

training dataset. At the same time, the testing is done on a random dataset collected from random users. The accuracy was 81% while testing the data. Xie et al., 2015, experimented with ensemble learning from facial images to predict a person's age. He formed team-based learning, which is correlated with a specific age group. ResNet is used for the implementation. For the x^{th} learner, the probability distribution is calculated using the following equation:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-x^2} dx &= \left[\int_{-\infty}^{\infty} e^{-x^2} dx \int_{-\infty}^{\infty} e^{-y^2} dy \right]^{(1/2)} \\ &= \left[\int_0^{2\pi} \int_0^{\infty} e^{-r^2} r dr d\theta \right]^{(1/2)} \\ &= \left[\pi \int_0^{\infty} e^{-u} du \right]^{(1/2)}. \end{aligned} \quad (4)$$

The probabilistic loss function is defined using equations (3) and (4) as follows:

$$\sum_{n=1}^n f^k = \frac{1}{N} \sum_{k=1}^k \sum_{i=0}^{n-1} \sum_{j=0} P_i^{kj} \log(p_i^k, j). \quad (5)$$

The system achieved an accuracy of 89.6%, with a loss of 2.9 at each level, which is very high for a CNN to consider it the best. The parameters such as tooth are also considered to predict the age by Kim et al. in 2020, using CNN while training the dataset using dental X-ray images. The first four molar images were captured using an X-ray. Using a heat map of the 16th, 26th, 36th, and 46th tooth is extracted to train the model to check the first molar presence. It was a pretty different trial of age prediction. Gradient weighted class activation mapping algorithm. The CNN of ResNet is used to train its own dataset of 2025 patients received by radiologists. The ages of only 20–29, 30–39, and 40–49 were used in the experimentation as tooth selection is mandatory to predict the age. The prediction accuracy was 89.05% for all three age groups. Age prediction using wrinkles present on the face was used by Sharma et al. in the year 2021 using CACD and UTK Face datasets with RGB images using FGNet CNN architecture to build a model. The model was executed for 200 epochs for the aging process. The author used the classes of 20–40, 41–60, and 60+ ages. Using the following equation, the output function is defined.

$$G(x) = \sum_{i=1}^{n-1} (C_y^f A_y^f) + x A_y^f, \quad (6)$$

where the first term represents the attention mask, x is the input image, and f is the respective class. The content mask is generated for each image, and the mouth portion is cropped. The error rate obtained was 0.001%, with an age prediction accuracy of 74. Using on UTKFace dataset, age prediction is experimented with by Sheoran et al., Sari et al., and Lobato et al., using deep CNN using transfer learning on the VGGFace2 pretrained model and ResNet50-f. The mean average error is calculated using the following equation:

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (7)$$

where y_i is the ground truth and \hat{y}_i is the predicted age of the i th sample. The mean error obtained was 6.098, and the accuracy was 5.91. We studied a survey on age prediction and tabularized in Table 1 the predictions of the literature about the findings. Most of the models use the face as a parameter to predict age. There were less number of articles that studied other parameters, such as predicting age using wrinkles and skin. These other medical articles [10–15] studied the prediction through a pupillometer. Gowroju et al. [16–18] proposed a method to segment the Pupil using a deep learning technique accurately.

The authors studied various age groups and concluded that pupil diameter changes as age grow up to the mid-40s and then the size of the Pupil decreases significantly as age increases [19]. The calculation also depends upon the luminance and refractive index of the eye. In the proposed system, we experimented on different aged people from benchmark datasets and a few randomly selected images as well as the live video to predict the age computationally to add it to the list of parameters to predict the age.

3. Proposed Method

Since the deep neural network has been proving its ability and intelligence in many feature extraction and segmentation applications, we considered one of the best performing architectures in the field of medicine was taken to implement the proposed method. We optimized the existing architecture of UNet and experimented with various optimizers and classifiers to finalize the proposed methodology. The overall block of the algorithm is listed in brief using 11 step process. The corresponding flowchart of the above algorithm is depicted in Figure 1.

3.1. Create a Trained Model. In this section, we described the datasets used for the experimentation and the pretrained models used for the efficiency calculation of the proposed method. In this section, we described the datasets used for the experimentation and the pretrained models used for the efficiency calculation of the proposed method. Deep Learning techniques may be taught more quickly by performing all controlling operations rather than sequentially. The proposed system used GPU for training the model. A GPU (Graphics Processing Unit) is a specific processor with dedicated secondary storage that performs the floating-point computations necessary for graphics translation. Regularization, such as dropout and early stopping of the GPU, helps increase the accuracy in the minimum number of epochs and decreases validation loss.

3.1.1. Datasets Used. The proposed system used three benchmark datasets: CASIA Ver-1, UBIRIS ver-2, and MMU. CASIA version 4 (Chinese Academy of Sciences' Institute of Automation) is one of the open databases containing 2639 infrared light images extracted from 239

TABLE 1: Pupil diameter according to the age group.

S. No	Age group	No of subjects	Diameter
1	15–19	12	2.0–2.1
2	20–29	15	2.3–3.5
3	30–39	23	4.1–5.12
4	40–49	23	5.21–6.15
5	50–59	21	4.81–5.77
6	60–69	12	3.58–4.63

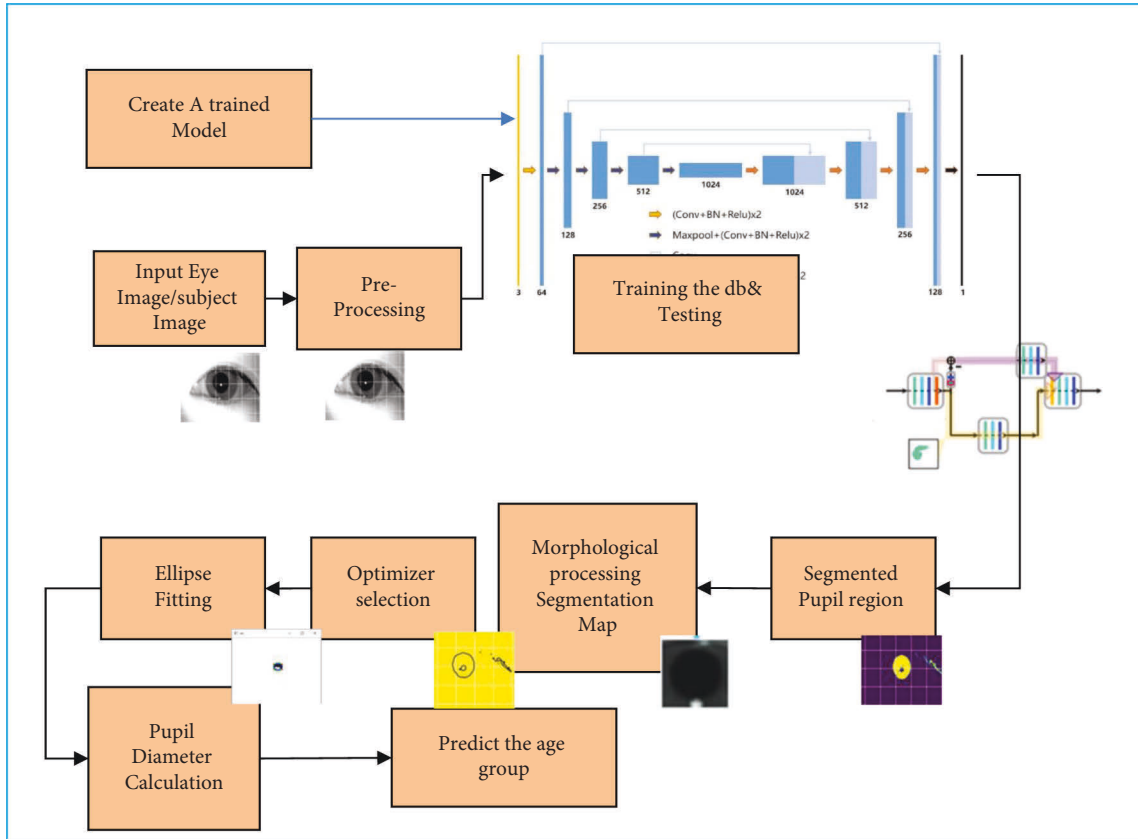


FIGURE 1: Flowchart of the proposed system.

subjects and eye images of different subjects of age groups 18–50 with a storage of 1.86 GB. UBIRIS is Noisy Visible Wavelength Iris Image Databases. The UBIRIS ver-2 is an open-source database that contains 10199 eye images collected from 241 subjects that contain detailed information on the eye, age, gender, and several other parameters. Multimedia University collects the MMU dataset for the iris biometric system, which contains 575 images with a high noise ratio. The three datasets were used individually for the proposed system to know the accuracy of the proposed system in different environments to predict the Pupil. The CASIA dataset is a noise dataset with infrared illumination, which makes automatic segmentation easier. UBIRIS is a colored dataset with significant occlusion and noise, and MMU is a collection of binary images with occlusion and noise. We experimented on three pretrained models to decide on the proposed UNet implementation. To experiment, we considered three CNN models, ResNet, UNet, and VGGNet.

3.1.2. Using ResNet 50. The deep neural network is hard to build because of the vanishing gradient problem where the Gradient is backpropagated in the layers, increasing the network's depth even more. As the ResNet is already built with 50 layers, as shown in Figure 2, the network's performance decreased numerously. In order to overcome this problem, we added skip connections in the network, which did not increase the network's performance, but the number of layers (depth) did not increase.

3.1.3. Using VGGNet. As the input to the VGG is fixed in size, i.e., of 224×224 image, the preprocessing of the image is mandatory since the image's input is 128×128 in the training dataset. The image is processed through the convolutional layers with a small receptive field of 3×3 , which makes the task of capturing Pupil very slowly. In the configuration of VGG, it also has the convolution filter of size 1×1 , where the pixel size is restricted to 1 in the spatial

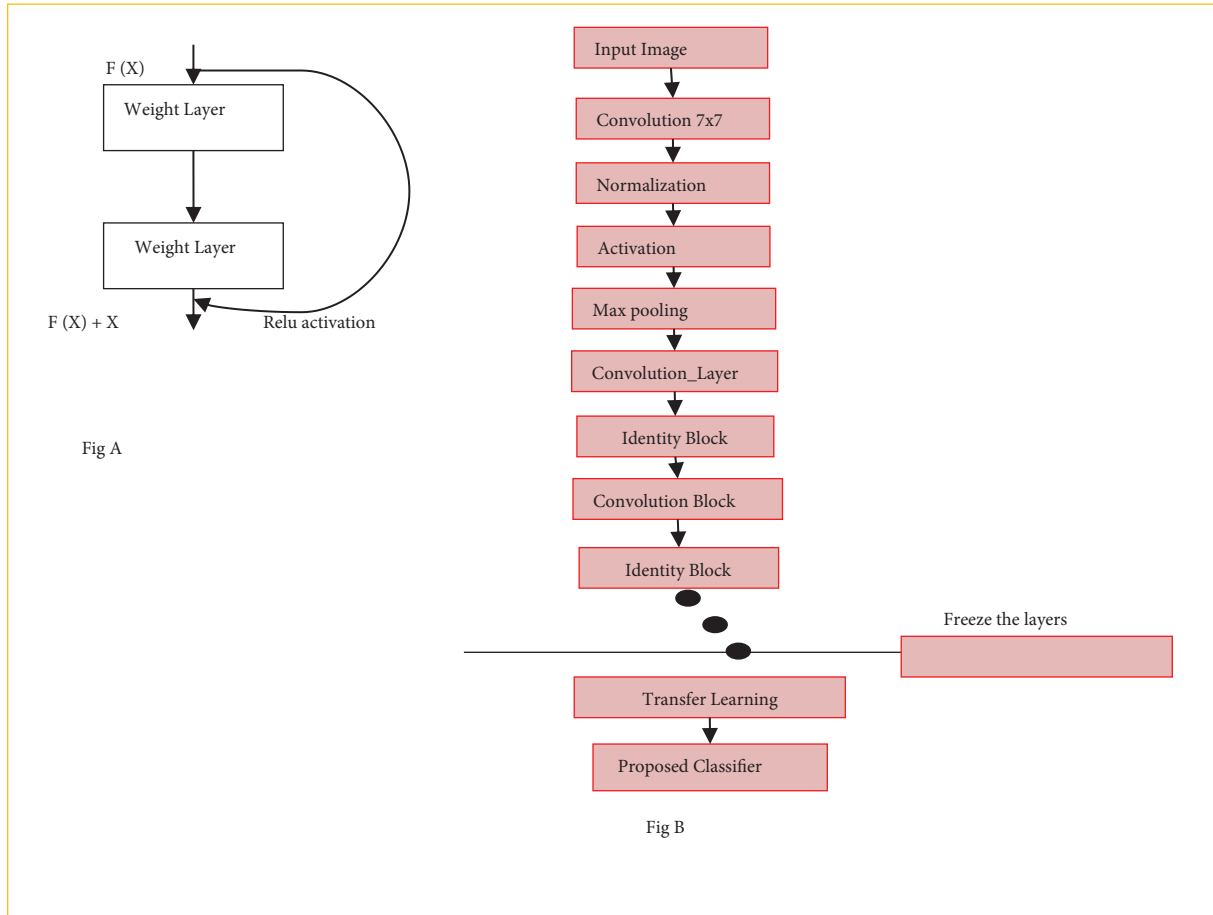


FIGURE 2: Resnet model.

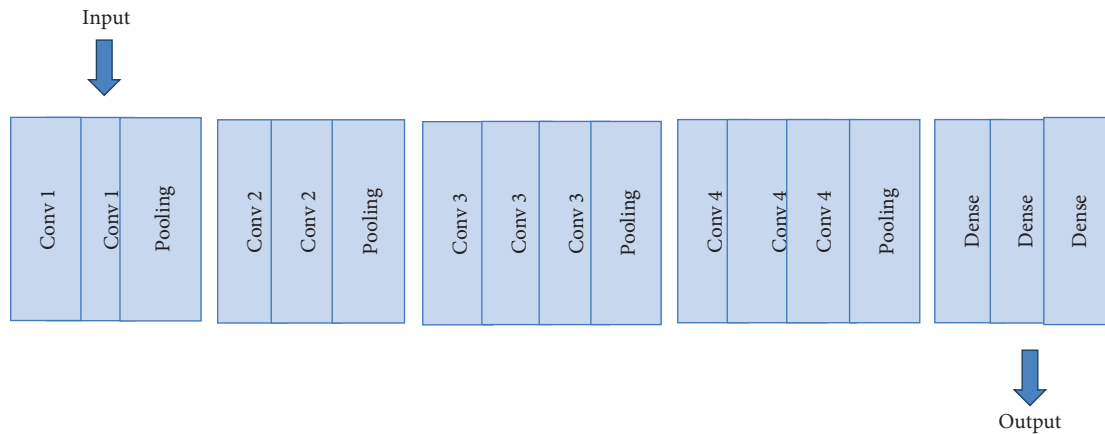


FIGURE 3: VGG architecture.

pooling. By the end of the architecture, due to the fully connected dense layers, the model is about to have 4096 channels where each fully connected layer classifies 1000 channels for the class, as shown in Figure 3. With the weight of 19 layers, it processes input with two conv3-64 layers and max-pooling between two conv3-128 layers, four conv3-256 layers, four conv3-512 layers, four conv3-512 layers, and three fully connected layers and a SoftMax layer with 144 number of parameters.

3.1.4. *Using UNet.* Next, we implemented UNet architecture for the same set of data. The UNet architecture is shown in Figure 4. The advantage of UNet is that it helps us to perform localization and classification together. The other existing models can also do the task, but due to overlapping redundancies, no accuracy in localization, and the need for a fully connected layer at the end of the connection, which leads to excessive network size, we opted UNet. The conventional UNet is modified and is used in the proposed

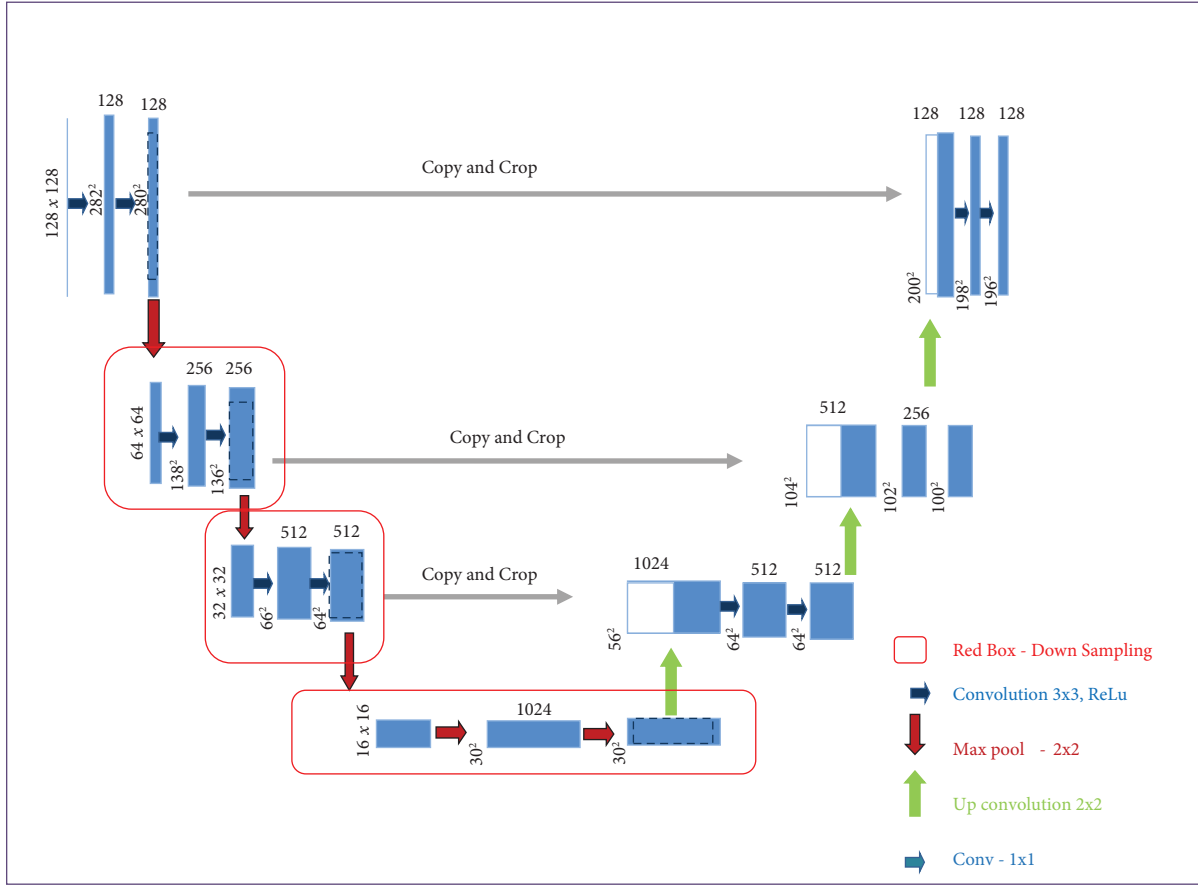


FIGURE 4: Proposed UNet model.

system. The edge mapping feature of the system uses equation (8). In the optimized model of UNet, the first two layers are used to collect the low-level features, and the last three are used for extracting high-level features.

$$L_e = - \sum_{i=1}^x \sum_{j=1}^y [G_e \log(s_e) + (1 - G_e) \log(1 - S_e)], \quad (8)$$

where (i, j) represents the pixel points from the predicted edge map s and G_e is the truth map, x, y are the width and height of the feature map, to extract the segmented part of the pupil. The sigmoid function activates the concatenation at the end to generate the segmentation feature map. The conventional UNet consists of five encoders, five decoder networks, and a fully connected SoftMax classifier. At each layer, while downsampling, it follows 3×3 convolutions and 2×2 ReLU function which is followed by max-pooling window with a stride value of 2. This downsampling extracts the features that double at each downsampling path. The low-resolution feature maps are replaced with high-level feature maps during the upsampling path while restoring the features to their original size to produce an output feature map. The concatenation is performed at each layer of encoding and decoding. This process is followed by ReLU activation to generate dense feature maps. After five consecutive encoding and decoding, the final fully connected SoftMax classifier classifies pixel by pixel classification. The conventional UNet

used in medical image processing to segment the tumor was a great attempt. However, this conventional UNet has a drawback regarding segmenting the Pupil from the eye. However, the conventional UNet suffers from 3 significant drawbacks when used with the pupil dataset. Firstly, UNet architecture eliminates redundancy by duplicating the low-resolution features to the next stage. This causes the edge information of the segmented object to be very rough and spiky. As in the proposed application, we need to find the edge of the Pupil accurately to determine the diameter. However, high-resolution edge data does not pass through the downsampling layers and does not contain enough information about those objects. A sample example is shown in Figure 5 because edge information is not appropriate to determine the Pupil for segmentation. The model is shown in Table 2: trains on a total of 1,928,322 trainable parameters that return the dimension of features of Pupil in the feature map. This feature map contains the noise and pupil information approximated with the edge map, as shown in Figure 6.

Hence, we modified conventional UNet to the proposed UNet model with Adaptive Movement Estimation optimization to perform segmentation.

3.2. Input an Image. Secondly, the input image is too small to get trained and generate feature maps. As the expected

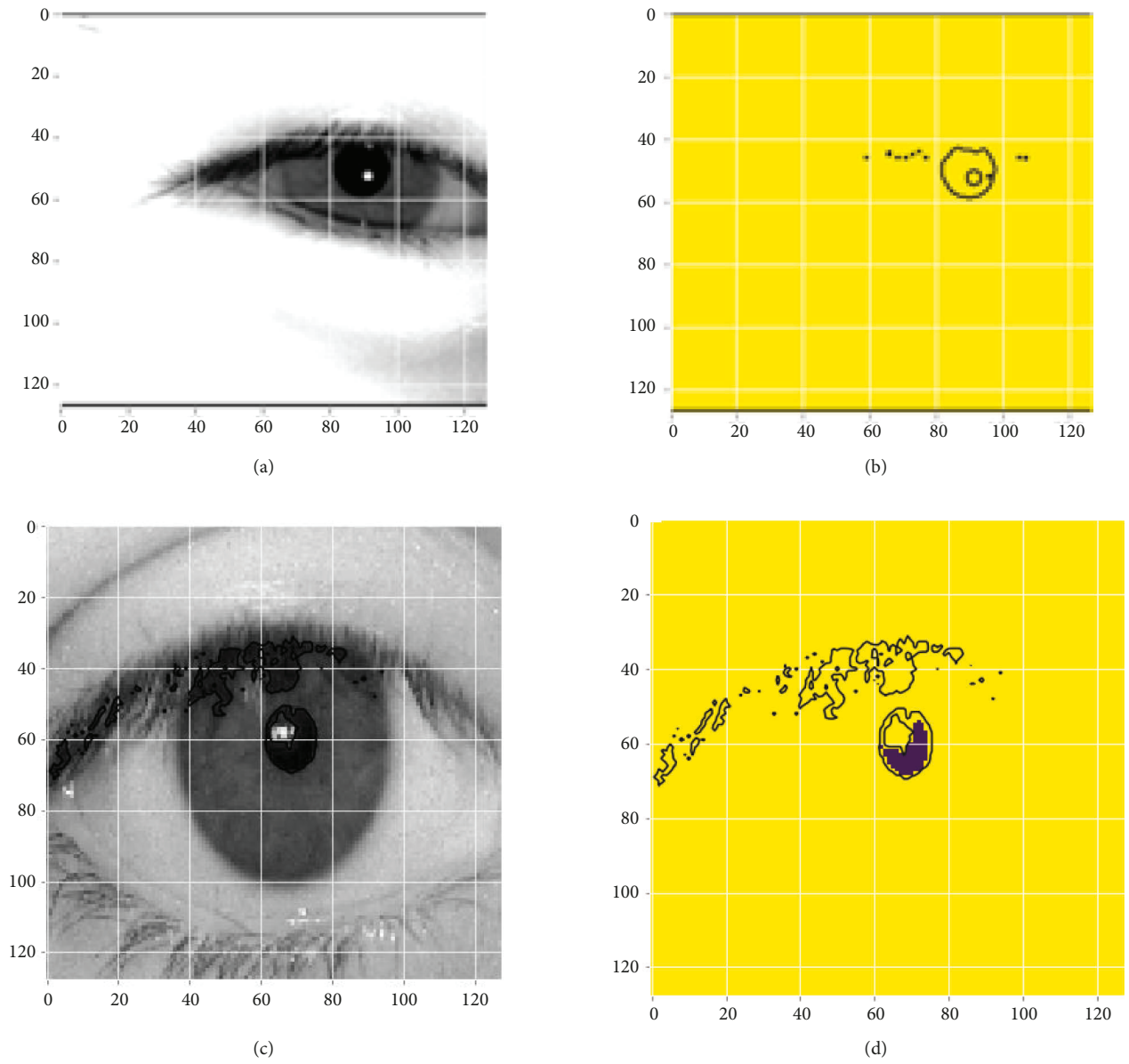


FIGURE 5: Feature map of Predicted Pupil along with noise information.

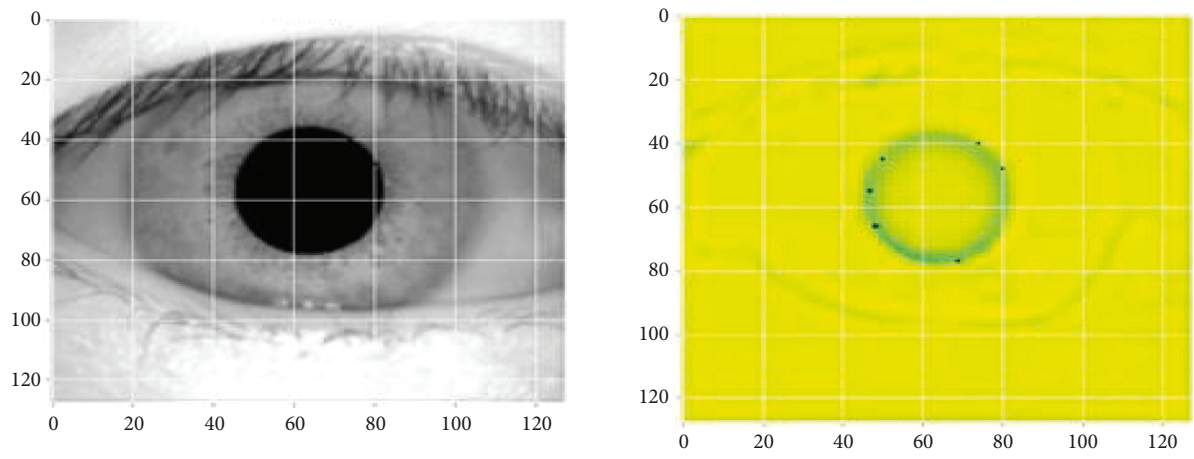


FIGURE 6: Edge Information extracted by Conventional UNet.

output is a single segmented pupil, and there is no chance of other outputs, the SoftMax used in the training model trains the model for a longer time. Instead, the proposed method uses a sigmoid function that works on two-class logistic regression to produce the feature map of the Pupil, followed by morphological thresholding to produce the accurate segmentation of the Pupil. Thirdly, when the input dataset is noisy, such as UBIRIS, the convolutional layers and the presence of max-pooling layers in the encoding path and decoding path increases the receptive field of the input image, with which the pixels around the Pupil are also classified as pupil area as the negative part in the feature map is set to 1. The sample images are shown in Figure 7, and this decreases the accuracy of segmentation of Pupil.

To address this situation, we trained the network with Leaky ReLU to preserve the pixels around the Pupil in a slow-decreasing part of the pupil area to prevent the dying ReLU problem caused by the second addressed problem caused by conventional UNet. Using the above steps, the UNet is reconstructed for the proposed system.

3.3. Preprocessing

3.3.1. Dataset and Training. The benchmark datasets CASIA and MMU were taken to train the model. For the segmentation, we considered the masks of each dataset to know the Pupil from each image. We divided the model into train and test folders. For each dataset, 80–20 ratio is considered for training-testing, respectively. The GPU-based system with an AMD processor is taken for training the model. We took 50 epochs and batch normalization while training the model. Sample images of each dataset are specified in Figure 8.

3.4. Training. We divided the data into training and testing sets. The image is converted into binary, and each dataset's masks are generated separately for training the model. While giving the input, each image is resized to 128×128 , as the convolution and max-pooling operations will generate the values of feature map parameters. The training is performed with a proper system configuration and generates a segmentation map. The following sections describe the system configuration and segmentation of Pupil in detail. Table 2 describes the summary of training proposed UNet. We cannot obtain a constant efficiency for the model because when the randomized state of the split changes, so does the model's accuracy. To prevent data leaking, keeping the testing data separate from the training data is best. The performance of the ML model must be assessed while it is being created using the training data. Here is when the value of cross-validation data becomes apparent. We split the data into training and validation sets with 80% and 20% images. The proposed system used a kernel (filter) size of 2×2 , batch size of 30, several epochs of 50, leaky relu as the activation, split ration as 80:20, and the learning rate as 0.001. Three optimizers were tested and checked the performance to use Adaptive moment estimation for training the model.

3.4.1. System Configuration. We used the GPU-based laptop hardware with 8 GB RAM, AMD processor and software configuration using Keras, Tensorflow, and OpenCV 3 using Python.

The datasets were considered according to the difficulty level of prediction, and the model is trained with ground truth pupil images of each dataset. The output is compared to the ground truth of the image to correct itself. We tested on three optimizers Gradient descent, RMS Prop, and Adaptive moment estimation. Based on the effectiveness shown by the Adaptive momentum algorithm while training, we chose this technique for the optimization in the proposed model. We evaluated training accuracy, recall, Precision, and F1 score for each trail we performed using the optimization. The hyperparameters are calculated using the confusion matrix defined as follows, and Table 3 shows the confusion matrix used in evaluating hyperparameters.

A positive label corresponds to the actual class of interest. Negative is anything apart from it. True positive is the corresponding accurate prediction from the collection of pixels. False negative corresponds to the other classes apart from the actual prediction class. Accuracy is calculated using the following equation:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}. \quad (9)$$

It is the fraction that measures the correctly classified pixels. The Precision can be calculated using the following equation:

$$\text{precision} = \frac{\text{tp}}{\text{tp} + \text{FP}}. \quad (10)$$

The approximate factual information can be calculated through Precision. It is the degree to which the class can be accurate. The recall is another hyperparameter that can be calculated through a confusion matrix using equation (11), which specifies the actual metric of several positives in the prediction:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (11)$$

Another parameter we need is the balancing between Precision and recall. It can be calculated using the F1-score of the model using the following equation:

$$f1\text{score} = \frac{2(P * R)}{P + R}, \quad (12)$$

where P is Precision and R is recall.

3.5. Segmentation of Pupil. The segmentation map is constructed from pooling operations in the encoding path followed by the skip connections on the input image. The process is shown in Figure 9. It describes the process of executing 2D. We start with a kernel and “stride” (slide) it over the 2D input data, multiplying the portion of the input it is currently on element-wise and then adding the results into a single output cell. Every area the kernel slides over is subjected to this same procedure, which creates a new 2D

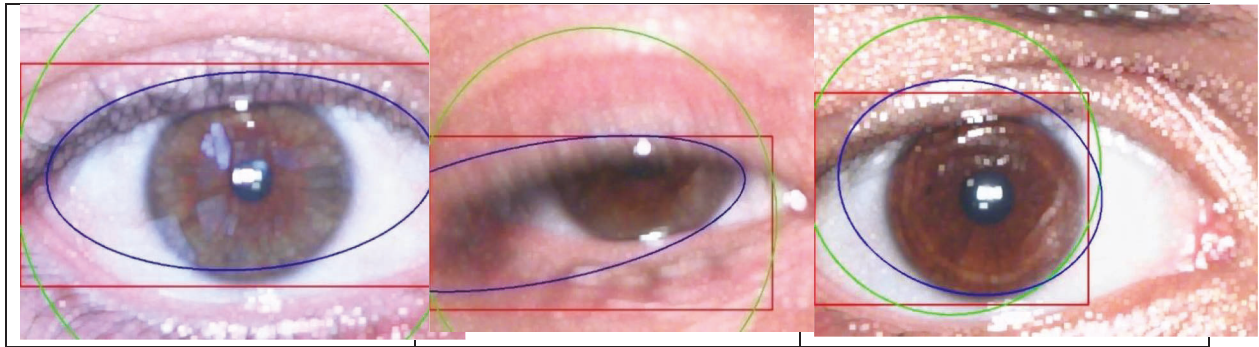


FIGURE 7: False nonpupil area detection in Noisy Images.

CASIA Version 1.4			
MMU			
UBIPeriocular			

FIGURE 8: Sample image from Datasets.

feature matrix. Stride describes the amount by which the kernel moves on the input feature matrix. In the animation below, the input matrix gets an additional stripe of zeros added from each of its four sides to guarantee that the output matrix is the same size as the input matrix. (Zero) padding is the term for this. The encoder module reduces the resolution by half for every pooling operation and increases the receptive field. The skip connections in the model reconstruct the original image to restore the changes in the image. Due to this, we generated the feature map, as shown in Figure 5. However, the feature maps that we obtained were noisy.

3.6. Morphological Processing and Segmentation Map. The main aim of the morphological operation is to reduce the

noise on the feature map. This process is described in Figure 10, and this process enhanced the image resolution by removing noise caused by eyelashes to give the feature map ready for ellipse fitting. We first applied a median filter to remove general noise caused, followed by erosion and dilation to enhance the pupil area. We applied thresholding on the enhanced pupil area to form a segmentation map to fit the ellipse.

3.7. Optimizer Selection. The main challenge in the machine learning algorithm is to minimize the loss at every epoch from the weights of each neuron. The optimizer helps to modify the weights and learning rate, which directly helps in reducing the overall loss. We experimented with three optimizers in the proposed method to predict the system's

TABLE 2: Training summary of UNet.

Layer (depth -index)	Output shape	Training parameters
Downsampling		
Conv2d: 3-1	32, 512, 512	320
BatchNorm2d: 3-3	32, 512, 512	64
Conv2d: 3-4	32, 512, 512	9,248
Conv2d: 3-8	64, 256, 256	18,496
BatchNorm2d: 3-10	64, 256, 256	128
Conv2d: 3-11	64, 256, 256	36,928
Conv2d: 3-15	128, 128, 128	73,856
BatchNorm2d: 3-17	128, 128, 128	256
Conv2d: 3-18	128, 128, 128	147,584
BatchNorm2d: 3-20	128, 128, 128	256
Conv2d: 3-22	256, 64, 64	295,168
BatchNorm2d: 3-24	256, 64, 64	512
Conv2d: 3-25	256, 64, 64	590,080
BatchNorm2d: 3-27	256, 64, 64	512
Upsampling		
ConvTranspose2d: 3-28	128, 128, 128	131,200
BatchNorm2d: 3-30	128, 128, 128	256
Conv2d: 3-32	128, 128, 128	295,040
BatchNorm2d: 3-34	128, 128, 128	256
Conv2d: 3-35	128, 128, 128	147,584
BatchNorm2d: 3-37	128, 128, 128	256
ConvTranspose2d: 3-38	64, 256, 256	32,832
BatchNorm2d: 3-40	64, 256, 256	128
Conv2d: 3-42	64, 256, 256	73,792
BatchNorm2d: 3-44	64, 256, 256	128
Conv2d: 3-45	64, 256, 256	36,928
BatchNorm2d: 3-47	64, 256, 256	128
ConvTranspose2d: 3-48	32, 512, 512	8,224
BatchNorm2d: 3-50	32, 512, 512	64
Conv2d: 3-52	32, 512, 512	18,464
BatchNorm2d: 3-54	32, 512, 512	64
Conv2d: 3-55	32, 512, 512	9,248
BatchNorm2d: 3-57	32, 512, 512	64

TABLE 3: Confusion matrix.

		Predicted class	
		Positive pred	Negative pred
True prediction class	Positive	TP (true positive)	FN (false negative)
	Negative	FP (false positive)	TN (true negative)

accuracy. The discussion and resultant loss values are explained in detail in the respective subsections.

3.7.1. Stochastic Gradient Descent Algorithm. This fundamental algorithm is commonly used in many deep learning techniques. It is the first-order derivative function to calculate the loss value. The weights of neurons are continuously altered so that they can reach minima. The Gradient is calculated using the following equation:

$$\text{for } i \text{ in range } (k): w_j := w_j - \alpha \frac{\partial J_i}{\partial w_j}, \quad (13)$$

where j is the cost of the i th training sample. Due to its nature of updating every time at every epoch, all parameters of the dataset were updated at every epoch, which increased

the time and loss. The corresponding result while training is shown in Table 4.

The loss using the SGD equation is calculated at every epoch; by the end of 40 epochs, the loss value is 2.36. The corresponding accuracy and loss graphs are shown in Figure 11. As we cannot directly test on the large dataset, we took the training sample of 700 images and their masks and a test sample of 140 images to perform the modeling.

The obtained loss value is 2.36. The confusion matrix can be drawn as shown in Figure 11. The hyperparameters for the SGD are shown in Table 5.

3.7.2. RMS Prop. Unlike the SGD, RMS Prop uses the decaying average of the gradients while adapting the parameter. It updates the weight for each input variable.

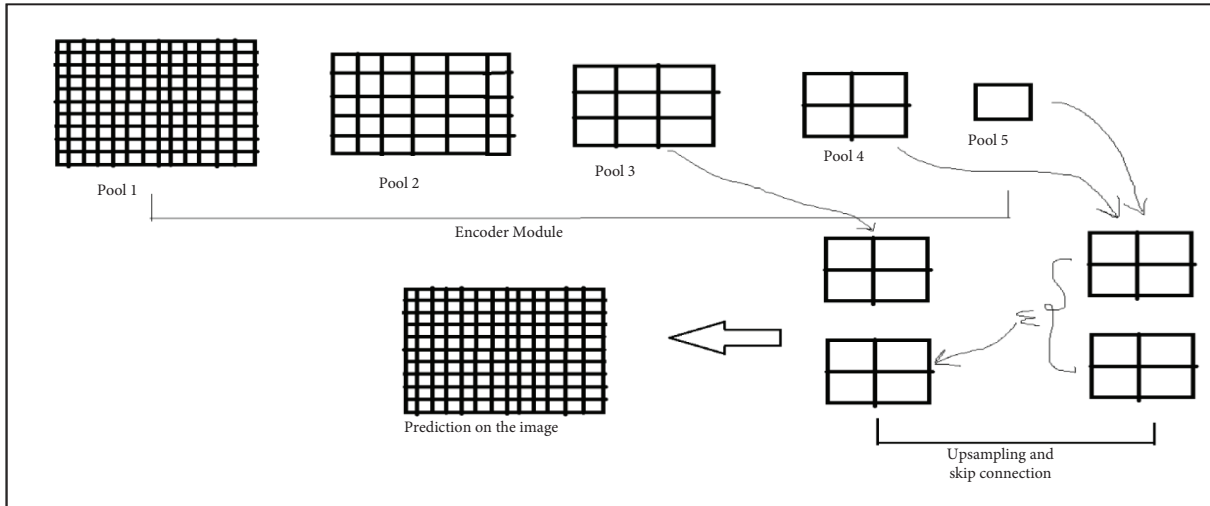


FIGURE 9: Segmentation feature map.

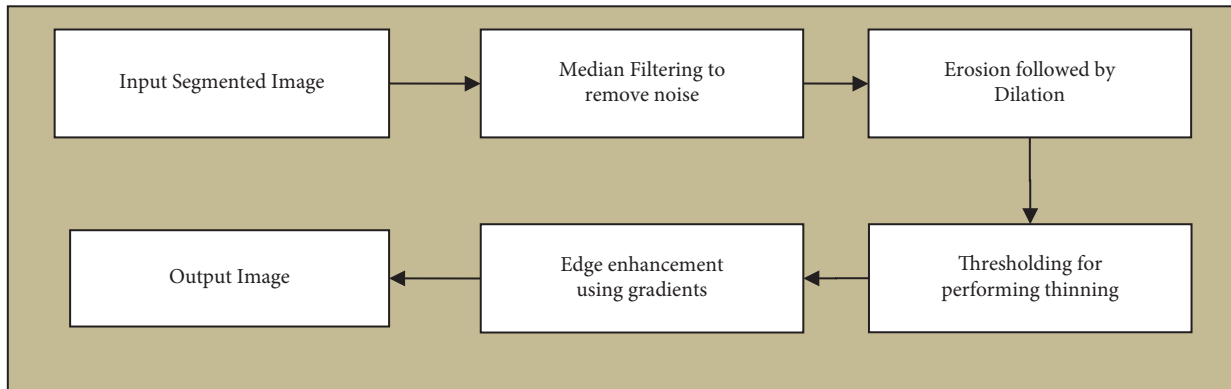


FIGURE 10: Flow chart for Morphological Processing.

TABLE 4: Training parameters using SGD in the model.

CNN layer	Operation	Output	Stride	Kernel	In	Out	Trainable variables
$64 \times 64 \times 32$	Convolution	$32 \times 32 \times 32$	2	2	32	32	0
$32 \times 32 \times 32$		$32 \times 32 \times 64$	2	3	32	64	0
$32 \times 32 \times 64$	Max pooling	$16 \times 16 \times 64$	2	3	64	64	9465
$16 \times 16 \times 64$	Conv	$16 \times 16 \times 128$	2	3	64	128	0
$16 \times 16 \times 128$	Conv	$8 \times 8 \times 128$	2	3	128	128	35642
$8 \times 8 \times 128$	Max pooling	$8 \times 8 \times 256$	2	3	128	256	32456
$8 \times 8 \times 256$	Conv	$16 \times 16 \times 128$	2	3	256	128	0
$16 \times 16 \times 128$	Conv	$16 \times 16 \times 256$	2	3	128	256	0
$16 \times 16 \times 256$	Conv	$16 \times 16 \times 128$	2	3	256	128	78954
$16 \times 16 \times 128$	Max pooling	$32 \times 32 \times 64$	2	3	128	64	123487
$32 \times 32 \times 64$	Conv	$32 \times 32 \times 128$	2	3	64	128	145863
$32 \times 32 \times 128$	Conv	$32 \times 32 \times 64$	2	3	128	64	754623
$32 \times 32 \times 64$	Conv	$64 \times 64 \times 32$	2	3	64	32	0
$64 \times 64 \times 32$	Max pooling	$64 \times 64 \times 64$	2	3	32	64	0
$64 \times 64 \times 64$	Conv	$64 \times 64 \times 32$	2	3	63	32	458692
$128 \times 128 \times 16$	Conv	$128 \times 128 \times 32$	2	3	16	32	459896
$128 \times 128 \times 32$	Conv	$128 \times 128 \times 16$	2	3	32	16	595656
$14 \times 14 \times 512$	Flatten	$7 \times 7 \times 512$	1	3	512	512	674569
$1 \times 1 \times 4096$	FC2	$1 \times 1 \times 1000$	7	2	4096	1000	745892

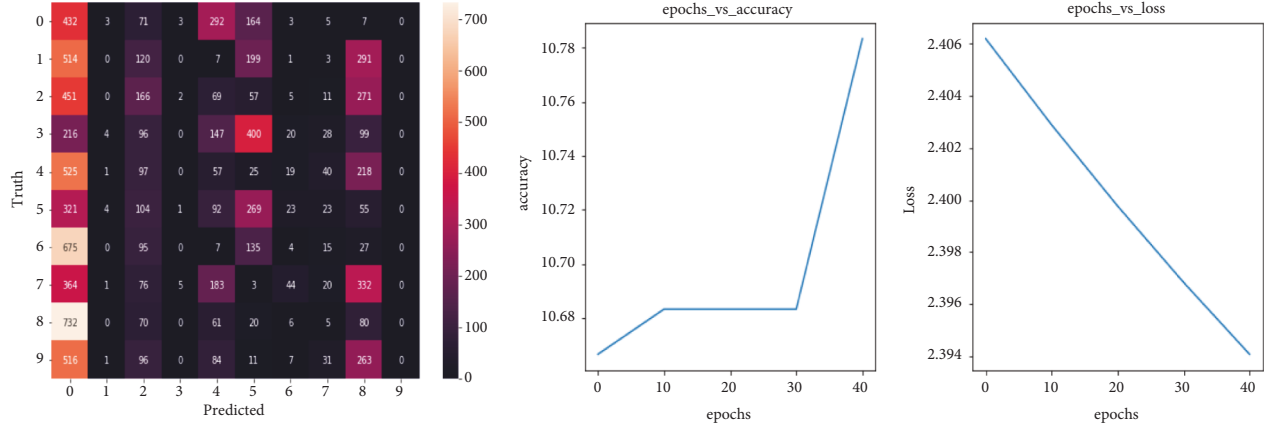


FIGURE 11: Confusion Matrix and Accuracy, Loss plots for SGD optimization.

TABLE 5: Hyperparameters of SG algorithm.

Stochastic gradient	
Precision	75.9
Specificity	74.2
F-score	76.2
Accuracy	71.9

Because of this property, we increased our learning rate with its beta value (Value of momentum) set to 0.9. The corresponding weights hence will be changed by the partial derivative functions equations (14) and (15) and weight w and gradient b can be calculated with equations (16) and (17).

$$v_{d_i} = \beta \cdot \partial_{d_i} + (1 - \beta) d_i^2, \quad (14)$$

$$v_{d_j} = \beta \cdot \partial_j + (1 - \beta) d_j^2, \quad (15)$$

$$w = w - \alpha \cdot \frac{d_i}{\sqrt{v_{d_i}} + \epsilon}, \quad (16)$$

$$b = b - \alpha \cdot \frac{d_i}{\sqrt{v_{d_j}} + \epsilon}. \quad (17)$$

For the proposed system, when the residual path carries the low-resolution value, v_{d_i} can be very low, with which the weight value may become zero. To prevent that, we are adding ϵ to prevent the denominator loss. The average of squared gradients balances the step size to avoid the vanishing gradient problem. The training using RMS Prop is shown in Table 6.

Because of the automatic adjusting of the learning rate, the trainable parameters increased more than the SGD implementation. Hence the loss value obtained is 2.30. The number of epochs is 50, the accuracy is 57% and the loss is 2.30. The confusion matrix and the learning curve are shown in Figure 12. The hyperparameters for the RMS Prop are shown in Table 7.

3.7.3. Adaptive Momentum Optimization. The adaptive moment optimizer uses squared gradients similar to the

RMS but calculates the moving averages. Due to the free selection of the learning rate for the neural network, it is easier to implement the technique. We took the initial moving averages in the proposed system as 0.9 and 0.999 for the minibatch and beta of the following equations:

$$m_i = \beta m_{i-1} + (1 - \beta) g_i, \quad (18)$$

$$v_i = \beta_2 v_{i-1} + l' - \beta_2 g_i^2, \quad (19)$$

where m and v the mean and variance of moving averages of minibatch and betas? These values express the first and second moments. The weights from the corrected moments can be calculated using the following equation:

$$w_t = w_{t-1} - \eta. \quad (20)$$

With the epsilon value as $1e-7$, the training values obtained for the proposed method are described in Table 8.

The loss and accuracy of the sample-tested data are shown in Figure 13. The loss value is 1.82, significantly less than the other two optimizers. The confusion matrix is represented in Figure 13.

The hyperparameters are calculated using the confusion matrix for Adaptive moment estimation and the values are tabularized as shown in Table 9.

When comparing the three optimizers, as shown in Table 10, plotted in Figure 14, we chose the adaptive momentum estimation as the suitable optimizer for the current dataset to perform the segmentation with minimum loss.

3.8. Ellipse Fitting. On the feature map obtained in the previous stage, we called a natural ellipse fitting algorithm using algebraic least squares using the following equation:

$$J(\theta, x_1, x_2, x_3 \dots x_1) = \sum_{n=1}^N \frac{\theta^i u(x_j) u(x_j)^i \theta}{\theta^2}, \quad (21)$$

where j is the Jacobian function, θ is the vectorized parameters, x is the vector variable, and $u(x_j)$ is the conic equation. The estimated pupil diameter can be corrected using the regression equation:

TABLE 6: Training parameters using RMS Prop in the model.

CNN layer	Operation	Output	Stride	Kernel	In	Out	Trainable variables
$64 \times 64 \times 32$	Convolution	$32 \times 32 \times 32$	2	2	32	32	0
$32 \times 32 \times 32$		$32 \times 32 \times 64$	2	3	32	64	1352
$32 \times 32 \times 64$	Max pooling	$16 \times 16 \times 64$	2	3	64	64	0
$16 \times 16 \times 64$	Conv	$16 \times 16 \times 128$	2	3	64	128	4256
$16 \times 16 \times 128$	Conv	$8 \times 8 \times 128$	2	3	128	128	45897
$8 \times 8 \times 128$	Max pooling	$8 \times 8 \times 256$	2	3	128	256	0
$8 \times 8 \times 256$	Conv	$16 \times 16 \times 128$	2	3	256	128	75686
$16 \times 16 \times 128$	Conv	$16 \times 16 \times 256$	2	3	128	256	77568
$16 \times 16 \times 256$	Conv	$16 \times 16 \times 128$	2	3	256	128	145689
$16 \times 16 \times 128$	Max pooling	$32 \times 32 \times 64$	2	3	128	64	0
$32 \times 32 \times 64$	Conv	$32 \times 32 \times 128$	2	3	64	128	214563
$32 \times 32 \times 128$	Conv	$32 \times 32 \times 64$	2	3	128	64	458796
$32 \times 32 \times 64$	Conv	$64 \times 64 \times 32$	2	3	64	32	8458923
$64 \times 64 \times 32$	Max pooling	$64 \times 64 \times 64$	2	3	32	64	0
$64 \times 64 \times 64$	Conv	$64 \times 64 \times 32$	2	3	63	32	7589666
$128 \times 128 \times 16$	Conv	$128 \times 128 \times 32$	2	3	16	32	47996669
$128 \times 128 \times 32$	Conv	$128 \times 128 \times 16$	2	3	32	16	47987642
$14 \times 14 \times 512$	Flatten	$7 \times 7 \times 512$	1	3	512	512	75852252
$1 \times 1 \times 4096$	FC2	$1 \times 1 \times 1000$	7	2	4096	1000	14256335

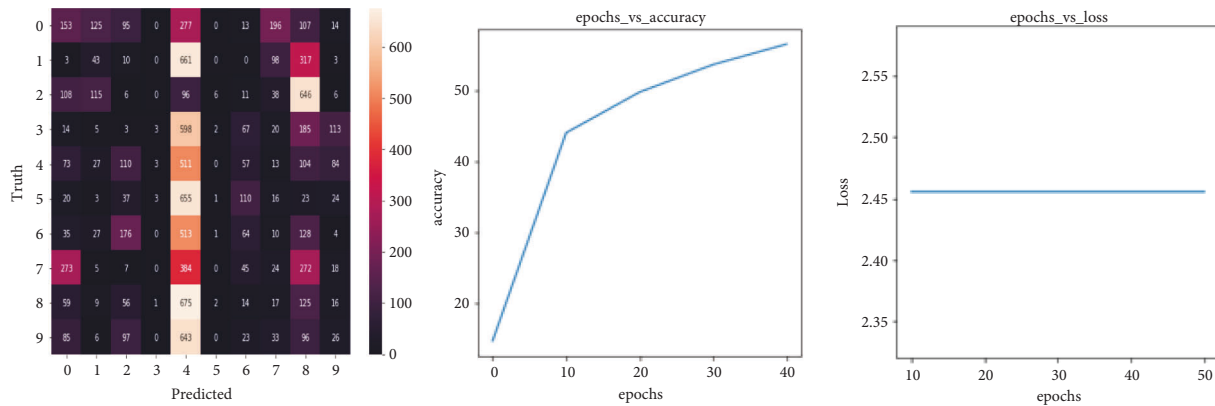


FIGURE 12: Confusion matrix and accuracy, loss plots for RMS prop.

$$E_c = P - k_0 - k_1x - k_2y, \quad (22)$$

where E_c is the corrected diameter, P is the predicted diameter values, x and y are horizontal and vertical gaze axes, and $k_{0,1,2}$ are regression parameters.

3.9. Age Prediction. In the last phase of the proposed methodology, we presented the age prediction using Figure 15, in which the calculated diameter is compared with specific values predicted using pupilometer to classify the age. Section 4.2 clearly shows the estimated Pupil and the age-predicted out of it.

4. Results & Discussion

Once the optimizer selection is finalized, we tried the proposed system’s segmentation, and it is accurate. We trained the system to segment the pupil part carefully to predict the diameter of the Pupil and the age. We considered three different step cases using the proposed method. First,

TABLE 7: Hyperparameters of RMS prop algorithm.

RMS prop	
Precision	75.3
Specificity	88.4
F-score	80.2
Accuracy	84.3

we analyzed the proposed system using the benchmark datasets, followed by checking the prediction on the image. Third, by using the live video. In these three cases, we calculated the accuracy for three cases by considering benchmark datasets from images and live videos. Three scenarios are explained in the following subsections.

4.1. Three Possible Cases of Input

4.1.1. Case I: Considering Benchmark Datasets for Segmentation. The benchmark datasets CASIA and MMU were considered to perform the segmentation of Pupil.

TABLE 8: Training parameters using adaptive moment optimizer prop in the model.

CNN layer	Operation	Output	Stride	Kernel	In	Out	Trainable variables
64 × 64 × 32	Convolution	32 × 32 × 32	2	2	32	32	0
32 × 32 × 32		32 × 32 × 64	2	3	32	64	0
32 × 32 × 64	Max pooling	16 × 16 × 64	2	3	64	64	9465
16 × 16 × 64	Conv	16 × 16 × 128	2	3	64	128	0
16 × 16 × 128	Conv	8 × 8 × 128	2	3	128	128	35642
8 × 8 × 128	Max pooling	8 × 8 × 256	2	3	128	256	32456
8 × 8 × 256	Conv	16 × 16 × 128	2	3	256	128	0
16 × 16 × 128	Conv	16 × 16 × 256	2	3	128	256	0
16 × 16 × 256	Conv	16 × 16 × 128	2	3	256	128	78954
16 × 16 × 128	Max pooling	32 × 32 × 64	2	3	128	64	123487
32 × 32 × 64	Conv	32 × 32 × 128	2	3	64	128	145863
32 × 32 × 128	Conv	32 × 32 × 64	2	3	128	64	754623
32 × 32 × 64	Conv	64 × 64 × 32	2	3	64	32	0
64 × 64 × 32	Max pooling	64 × 64 × 64	2	3	32	64	0
64 × 64 × 64	Conv	64 × 64 × 32	2	3	63	32	458692
128 × 128 × 16	Conv	128 × 128 × 32	2	3	16	32	459896
128 × 128 × 32	Conv	128 × 128 × 16	2	3	32	16	595656
14 × 14 × 512	Flatten	7 × 7 × 512	1	3	512	512	652342
1 × 1 × 4096	FC2	1 × 1 × 1000	7	2	4096	1000	687519

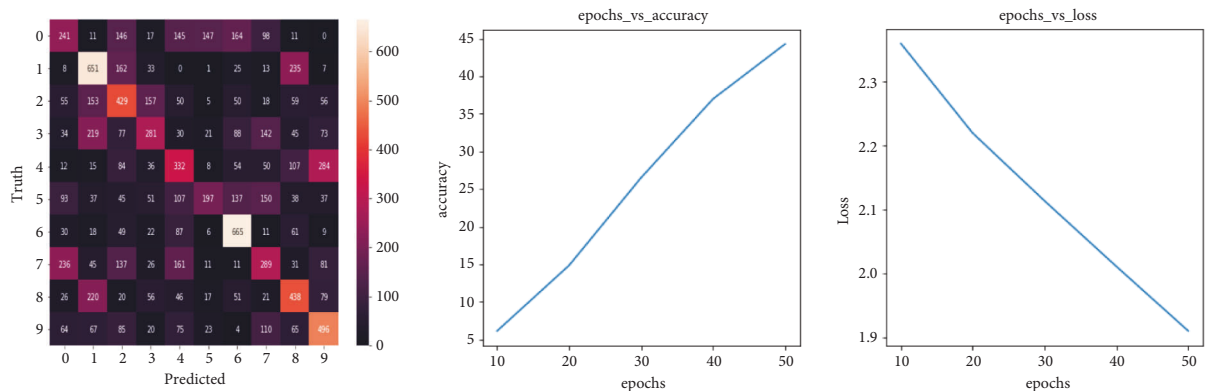


FIGURE 13: Confusion matrix and accuracy, loss plots for adaptive moment optimizer.

TABLE 9: Hyperparameters of adaptive moment algorithm.

Adaptive movement estimation	
Precision	93.0
Specificity	96.4
F-score	94.7
Accuracy	96.0

TABLE 10: Comparison graph of optimizers.

	Stochastic gradient	Root mean square prop	Adaptive movement estimation
Precision	75.9	75.3	93.0
Specificity	74.2	88.4	96.4
F-score	76.2	80.2	94.7
Accuracy	71.9	84.3	96.0

Figure 16 shows the segmented part of the Pupil after fitting an ellipse to it.

The original input images are taken from six subjects in the CASIA database of different age groups. The low-intensity values from Figure 16 are carefully discarded due to

the morphological processing after the feature map generation. The edges of subjects are also carefully taken without any further loss of resolution due to the residual path we added to the architecture. CASIA is no noise database, as it is captured through infrared near the camera.

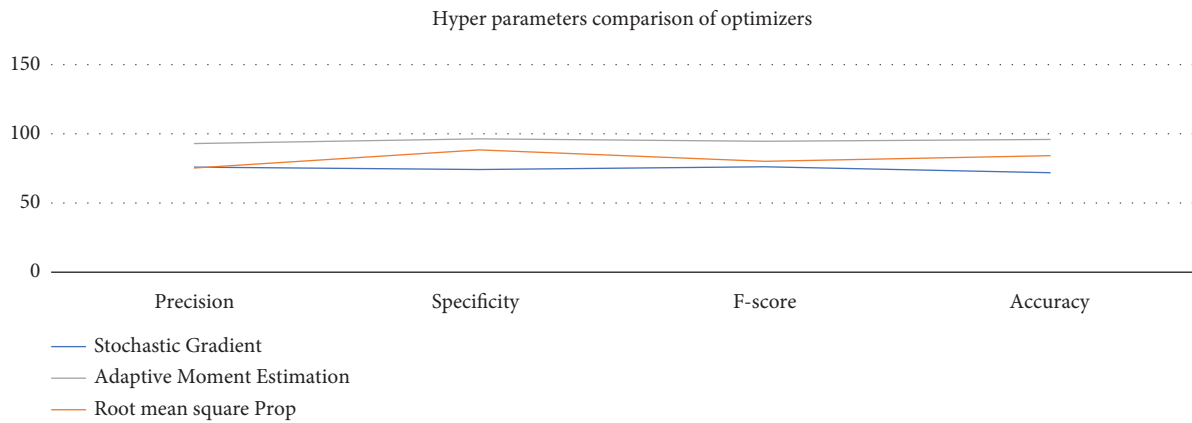


FIGURE 14: Graph of hyperparameters.

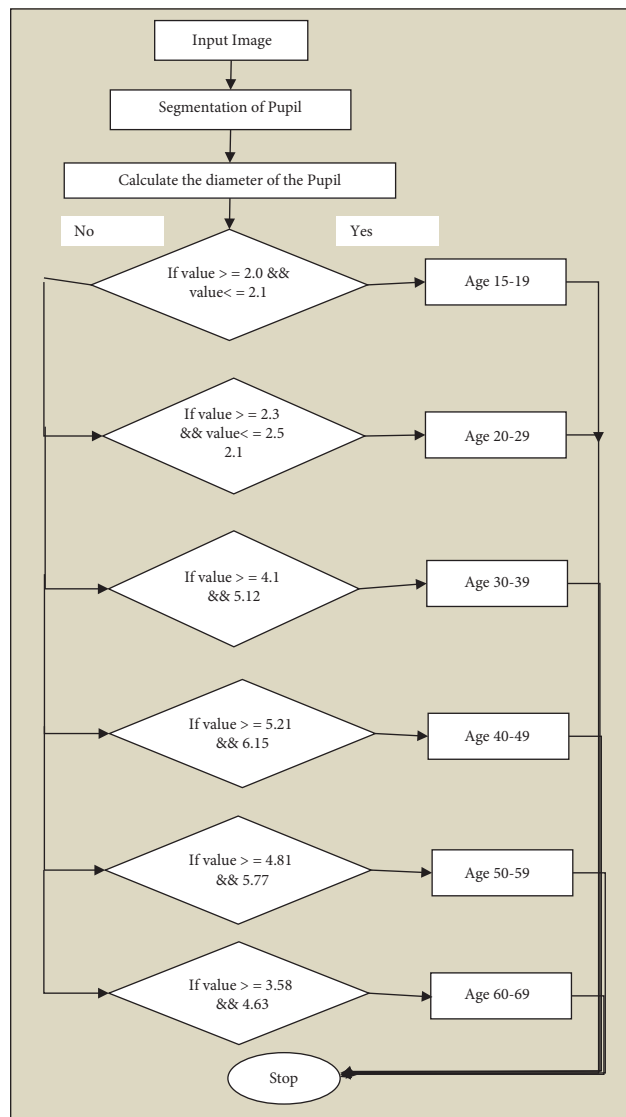


FIGURE 15: Flowchart of age classification.

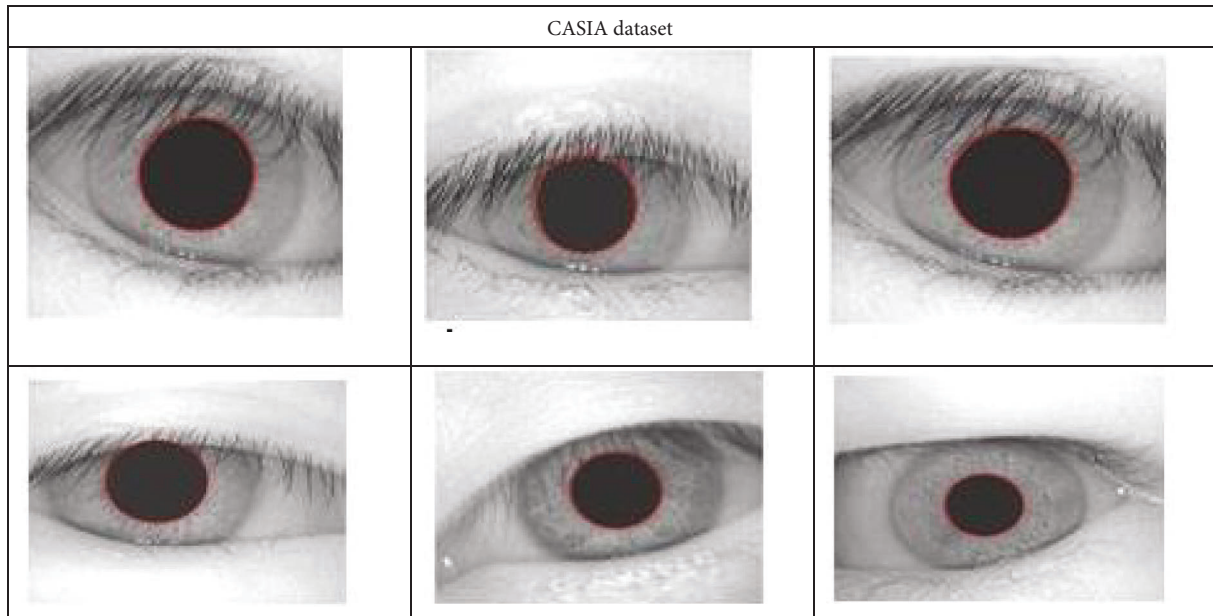


FIGURE 16: Ellipse fitting on CASIA dataset.

The segmentation of Pupil is almost accurate to 100% due to significantly less noise. The same experimentation is performed on the MMU dataset and UBIRIS also. The resultant fitted images are shown in Figure 17. Due to the occlusion and noise present in the images, the prediction was not 100% accurate. However, the accuracy we obtained for the CASIA dataset is 98%. For the MMU dataset, it is 94%, and for UBIRIS, it is 91%. By understanding the results obtained with the benchmark datasets, we designed the model to suit live video. There were three challenges we needed to face for live video. Those are as follows: .

- (1) A video is a moving object. Performing segmentation needs a static image.
- (2) We will receive a full image of the Subject. We directly cannot see the eye image of the person.
- (3) The clarity of the input video is random, which depends upon the background and lighting conditions of the room. Because of these three challenges, we first considered segmentation on the image, which aided in proceeding to the live video. We discussed the result analysis of the image in the following case. The model accuracy and loss graphs are shown in Figure 18. The proposed implementation is then tested on benchmark datasets. Surprisingly the output was at most accurate on the noisy dataset, as shown in Figure 19. Hence the proposed system then trained on the input eye without segmentation by generating the mask to identify the pupil part. In case II, we explained the corresponding scenario.

4.1.2. Case II: Confirming Age Prediction from the Image. For this case, we considered the dataset of UBIRIS and periocular images to localize the pupil area. As the input

image is noisy with other features such as skin eyebrows, segmentation of the Pupil is a complex task. Hence, we generated the eye masks to train the network for pupil segmentation. The experiment was quite working and the segmentation was successful for all the images. The resultant prediction is shown in Figure 20. With this advancement, we moved on to the segmentation of live video. The analysis is discussed in the following case.

4.1.3. Case III: Segmentation on Live Video. We considered the specification of the system as an ASUS gaming laptop, with a GPU-based processor and 8 GB of Ram capacity to run the algorithm. The localization process is a bit difficult in the case of live video. In the proposed algorithm, we performed preprocessing to remove the noise from the full facial image to generate pixel data of the eye. The preprocessing steps are as follows.

Step 1. Predict the face of the Subject.

Step 2. Locate facial points using landmarks detection.

Step 3. Localize the eye portion out of the region.

Step 4. Load the model to predict the Pupil.

Step 5. Segment the pupil part.

Step 6. Fit the ellipse.

We explained each step in detail in further discussion. In Step 1, we used the voila Jones face detection algorithm, which used the Haar cascade frontal face classifier to detect the face's x , y , w , and h parameters to define the rounded box using ROI selection. As the viola Jones algorithm suffers only

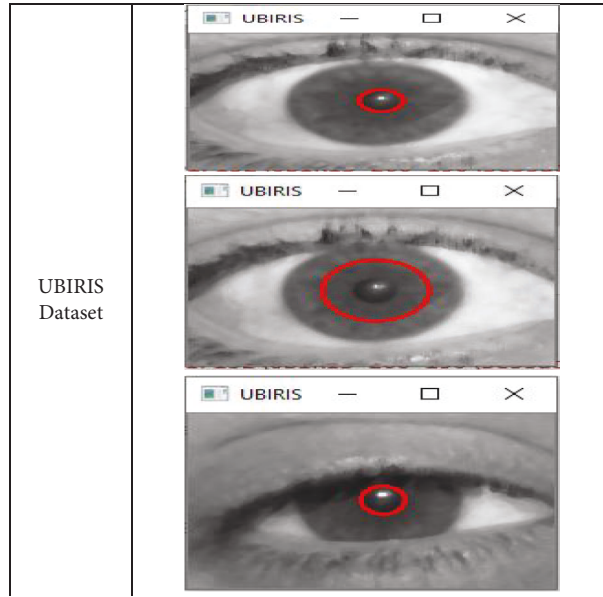


FIGURE 17: Result of segmentation on MMU and UBIRS datasets.

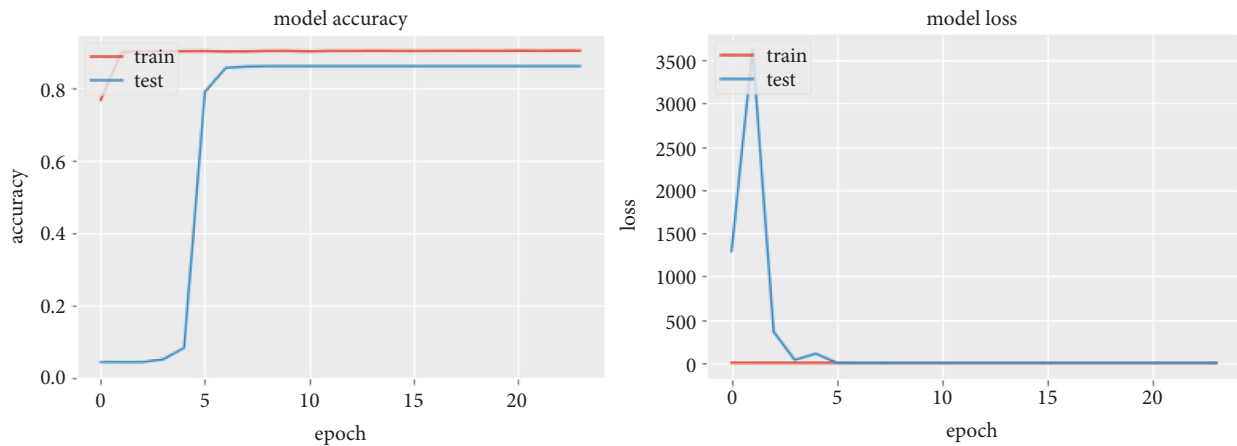


FIGURE 18: Plots are drawn against accuracy and loss.

binary classification and is too sensitive for low and high brightness and high false detection rates in further classification of facial parts, we did not consider this for eye localization. The resultant image after facial detection s is shown in Figure 21.

Once we obtain the face from the input video, the top part of the work is done. Then we proceeded to the next step, locating facial points using 49 landmarks. We checked the various landmark detection algorithms from the literature and selected the best detection using 49 landmarks to predict various parts of the face, as shown in Figure 22.

The 49 landmarks estimation algorithm is processed to produce the feature map of the proposed system. The corresponding feature map is shown in Figure 23(a). We figured out the eye position to perform the segmentation. In Step 3, the Eye position is localized, as shown in Figure 23(b). The proposed method is pupil segmentation. As the clarity is less in the live video, we implemented thresholding to perform the

segmentation. As we implemented thresholding on the binary image, as proposed, we could analyze the image quickly. The resultant segmentation is shown in Figures 23(c) & 23(d).

The pupil positions were extracted from the video using key points. Sample key point selection from the console is shown in Figure 24. The key points help us to locate the ellipse out of the face. In Step 5, we called the ellipse fitting algorithm to fit the ellipse. The segmented part of the Pupil is further processed from the image to perform age prediction. The diameter of the Pupil is shown in Figure 25.

It is calculated to check the age of the person. From the literature as we reviewed several articles written by researchers to check the pupil diameter variation according to age, we confirmed our research from various authors as shown in the literature. The diameter of the Pupil is calculated from the segmented Pupil to predict the subject's age. We experimented on random images collected from the Internet on live video.

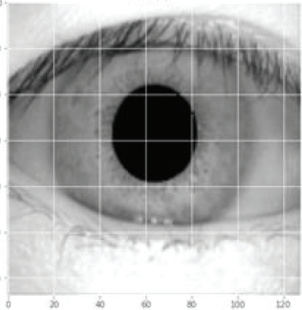
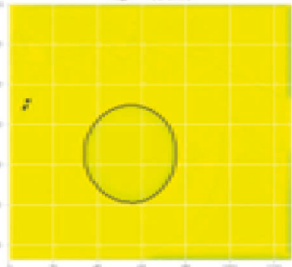
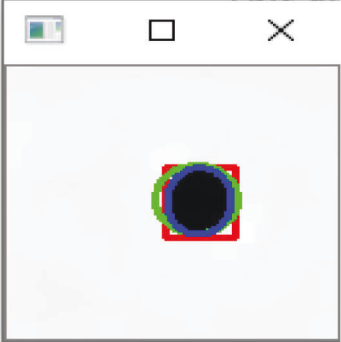
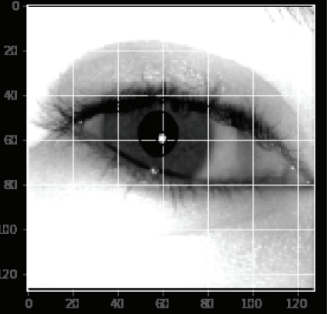
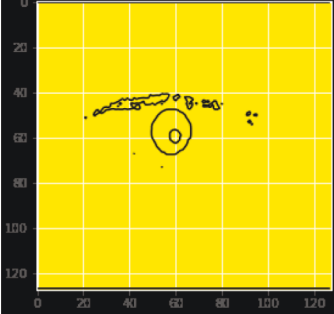
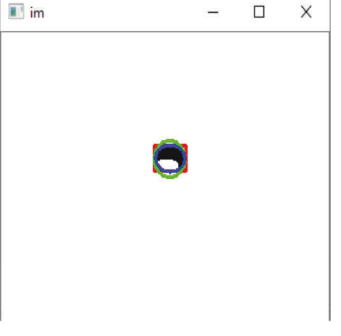
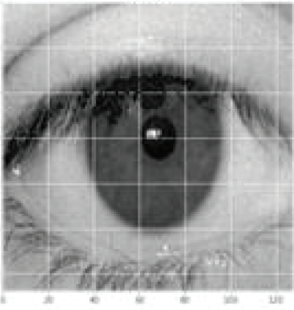
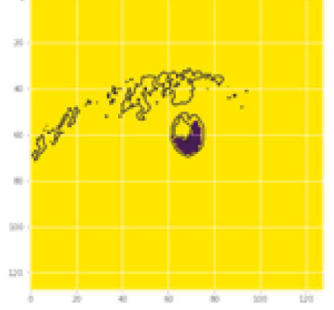
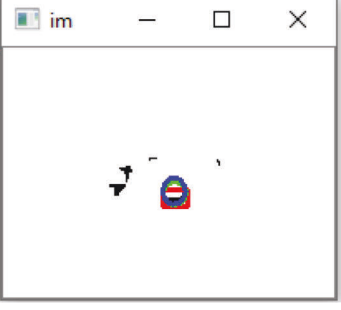
Dataset	Input Image	Segmented Map	Pupil Prediction
CASSI A V 1.0			
MMU			
UBIRIS Version-2			

FIGURE 19: Pupil segmentation and fitting the ellipse on segmented pupil on standard datasets, CASIA and MMU.

4.2. Age Prediction. We took sample images from various subjects on the Internet with different age groups to analyze the proposed system. We tabularized different people's ages and pupil diameters using a pupillometer. We calculated the Pupil's diameter from the ellipse obtained in the segmentation process. The estimated pupil sizes of different age groups are listed in Table 11. Using the mean data obtained from various kinds of literature, we predicted the diameter of the Pupil: Using the above analysis, the person's age group is predicted. The accuracy of the proposed system is 97.4%. We considered 7 age groups, i.e., '1-2', '3-9', '10-20', '21-27', '28-45', '46-65', and '66-100'. The corresponding ages we considered for the experimentation were tested under the proposed system. The calculated ages and pupil diameters are shown

in Table 11. The resultant image after calculating the age is shown in Figure 24.

Figures 24(a) and 24(b) explain the correctness of predicting age. Figures 24(b) and 24(c) show the incorrect prediction due to an unidentified Pupil. We tested nearly 80 images using the proposed system. The graph explains that the pupil diameter is in specific diameters for different age groups. There is a clash in identifying the pupil diameter calculated for the age group 30 to 60 with individual age groups of 28-45. We are trying to find a way to eliminate this redundancy using another parameter to calculate the age.

We considered the dataset of 800 subjects of different age groups. We divided them into six groups for easier calculation. The groups are grp 0: '1-9', grp 1: '10-20', grp 2:

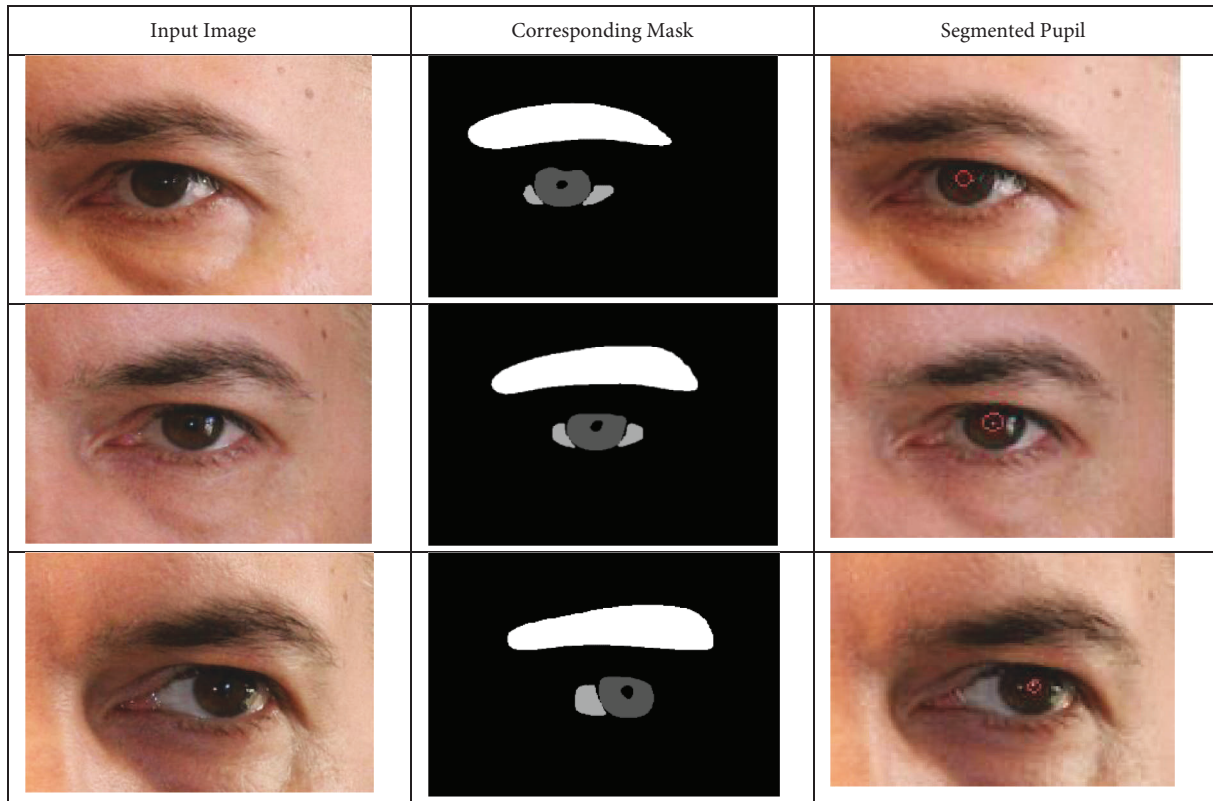


FIGURE 20: Input image from UBIRIS dataset and its corresponding mask.

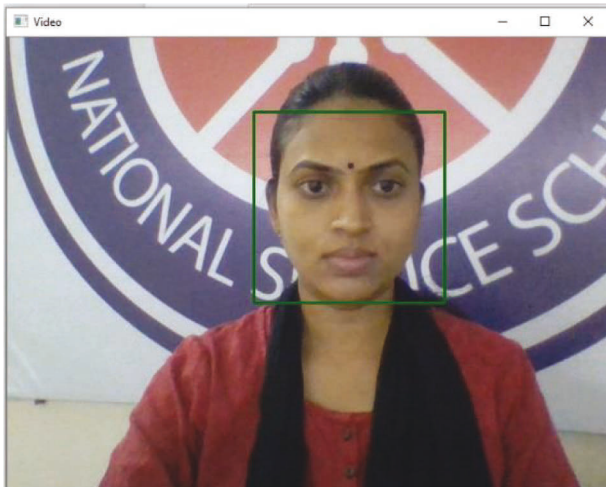


FIGURE 21: Input video from webcam—detecting face using classifier.

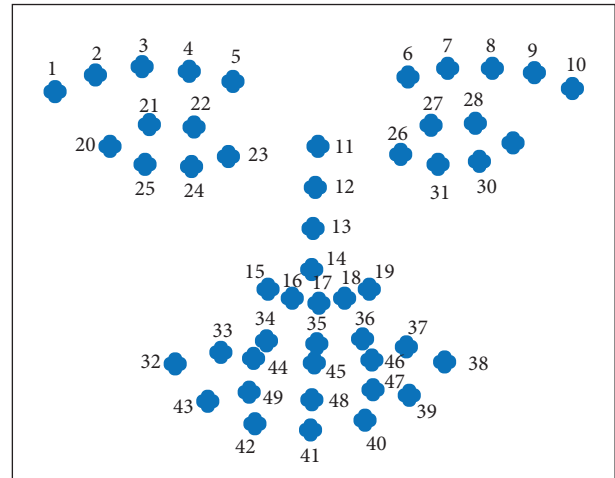


FIGURE 22: 49 landmarks to detect the eye portions of the Subject.

'21–27', grp 3: '28–45', grp 4: '46–65', grp 5: '66–100'. However, we did not consider the grp 0 for the experimentation. Few images were collected from the UBIRIS periocular dataset. Moreover, few were collected on their own at university premises and from the Internet. For various age groups, the diameter is recorded and plotted in Figure 26. The graph of the entire testing set database is

shown in Figure 27. The diameter values are different for different age groups and can be identified here. The hyperparameters of the proposed system were calculated and shown in Table 12.

5. Analysis of Hyperparameters

The proposed system has experimented on four different datasets, three benchmark datasets CASIA, UBIRIS, MMU,

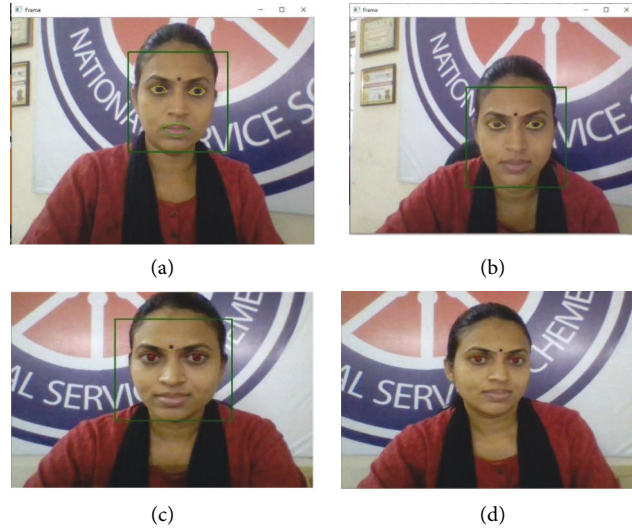


FIGURE 23: (a) Detected landmarks 23. (b) Localizing the eye portion. (c & d) Resultant segmentation of Pupil.

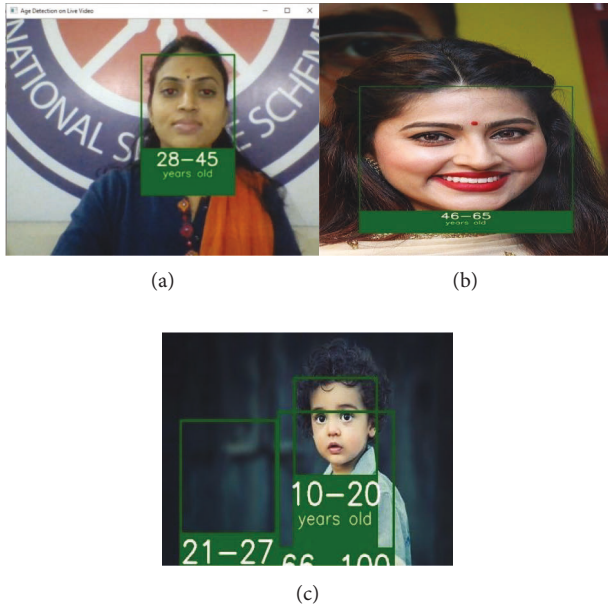


FIGURE 24: (a) Input live video-resultant age group. (b & c) Random image from the Internet.

and Own dataset. Table 12 shows the accuracy obtained on each dataset. The benchmark datasets have achieved the accuracy of 89.6, 95.3, 93.1, and 95.62, respectively. The accuracy for each dataset varied according to the occlusion present in the data.

6. Comparison

In this section, we compared the proposed system accuracy with existing systems, and the graphs were drawn on the comparison in the subsequent section. As stated in the literature, several other parameters predict a person's age. We calculated the mean error using the following equation:



FIGURE 25: Diameter of pupil.

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (23)$$

where y_i is the ground truth and \hat{y}_i is the predicted age of the i th sample. The error obtained by various authors is collected and compared with the proposed method to check the accuracy. It is shown that the error rate is more for small kids of age 0–9. As the age grows, the error value is decreased and reaches the minimum value. The error rate increased slowly up to 3.08 for ages above 70.

Comparatively, the proposed system has shown better accuracy than the existing systems. The accuracy of the existing systems is also compared and shown in Table 13.

Table 14 shows the proposed system's accuracy over other state-of-the-art methods. The proposed system shows better accuracy over existing systems and prediction parameters.

TABLE 11: Predicted diameter values of pupil.

Sl. No	Age of the subject	Width and height of the ellipse	Diameter calculated by the proposed method
1	24	63.93853759765625 65.2686996459961	2.765833
2	27	68.04349517822266 69.22743225097656	3.030417
3	27	61.34874725341797 65.39747619628906	2.765833
4	27	62.53804016113281 65.19945526123047	2.765833
5	20	57.815773010253906 61.56900405883789	2.236667
6	24	87.5483169555664 90.62991333007812	5.67625
7	35	83.653076171875 85.93659973144531	5.411667

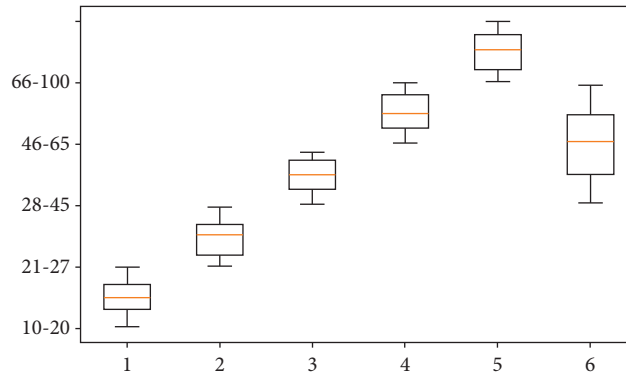


FIGURE 26: Plot representing pupil measurements and corresponding age group distribution.

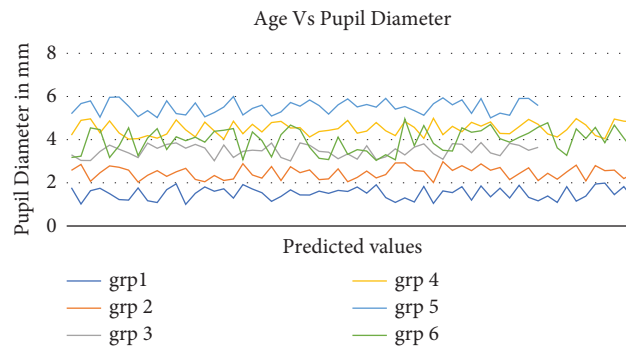


FIGURE 27: Graph of Predicted values from the proposed system.

TABLE 12: Hyperparameters of the proposed system.

Dataset	Age grp	Size of the data	Rate of accuracy	Overall classification rate
Test data	'10-20'	60	80%	88.6%
	'21-27'	50	86%	
	'28-45'	100	98%	
	'46-65'	50	82%	
	'66-100'	50	97%	
Training data	'10-20'	20	94.7	89.6
	'21-27'	20	88.5	
	'28-45'	24	89.6	
	'46-65'	36	84.5	
	'66-100'	14	83.6	

TABLE 12: Continued.

Dataset	Age grp	Size of the data	Rate of accuracy	Overall classification rate
UBER IS	'10-20'	30	89.4	95.3
	'21-27'	25	98.3	
	'28-45'	21	97.3	
	'46-65'	35	96.8	
	'66-100'	29	94.9	
MMU	'10-20'	10	87.6	93.1
	'21-27'	10	89.3	
	'28-45'	20	97.4	
	'46-65'	20	95.4	
	'66-100'	25	96.1	
Random dataset	'10-20'	60	97.5	95.62
	'21-27'	50	89.3	
	'28-45'	100	95.3	
	'46-65'	50	98.4	
	'66-100'	60	97.6	
Avg				93.40

TABLE 13: Comparison of Accuracy with state-of-the-art methods.

	[3]	[5]	[4]	[6]	[7]	[9]	Ours
Accuracy (%)	78.9	79	81	89.6	89.05	74	93.40

TABLE 14: Comparison of various parameters and their mean error rate using various parameters.

Age group	Using other parameters			Using face parameter					
	Ours using pupil parameter	Using molar parameter [3]	Using wrinkles as a parameter [5]	[4]	[6]	[7]	[9]	[11]	
0-9	6.84	11.21	6.33	11.97	14.33	9.86	13.63	14.08	
10-19	3.46	6.23	4.32	9.85	8.78	7.75	10.45	9.56	
20-29	3.36	7.95	6.31	9.78	9.96	8.42	10.45	9.74	
30-39	2.22	8.17	7.3	9.69	9.25	7.12	9.45	8.12	
40-49	1.71	8.64	1.23	10.35	8.75	8.56	11.63	9.79	
50-59	2.35	9.43	2.31	10.56	9.63	11.55	13.75	10.63	
60-69	2.82	11.12	6.32	13.78	10.12	12.78	15.96	13.85	
70-89	3.08	15.56	8.12	19.96	16.45	16.96	20.13	19.68	
Total	3.23	9.78875	7.37	11.9925	10.90875	10.375	13.18125	11.93125	

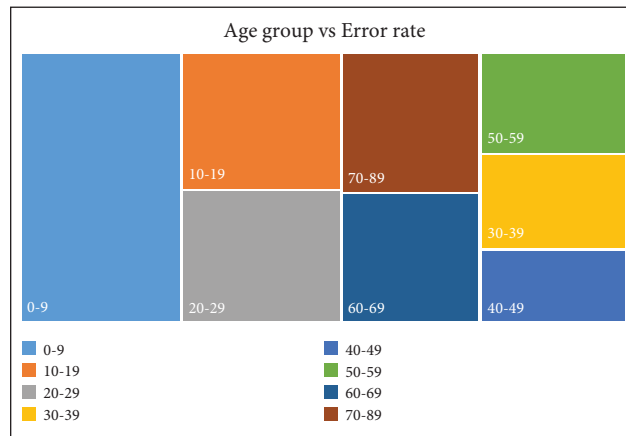


FIGURE 28: Age group and the corresponding error rate.

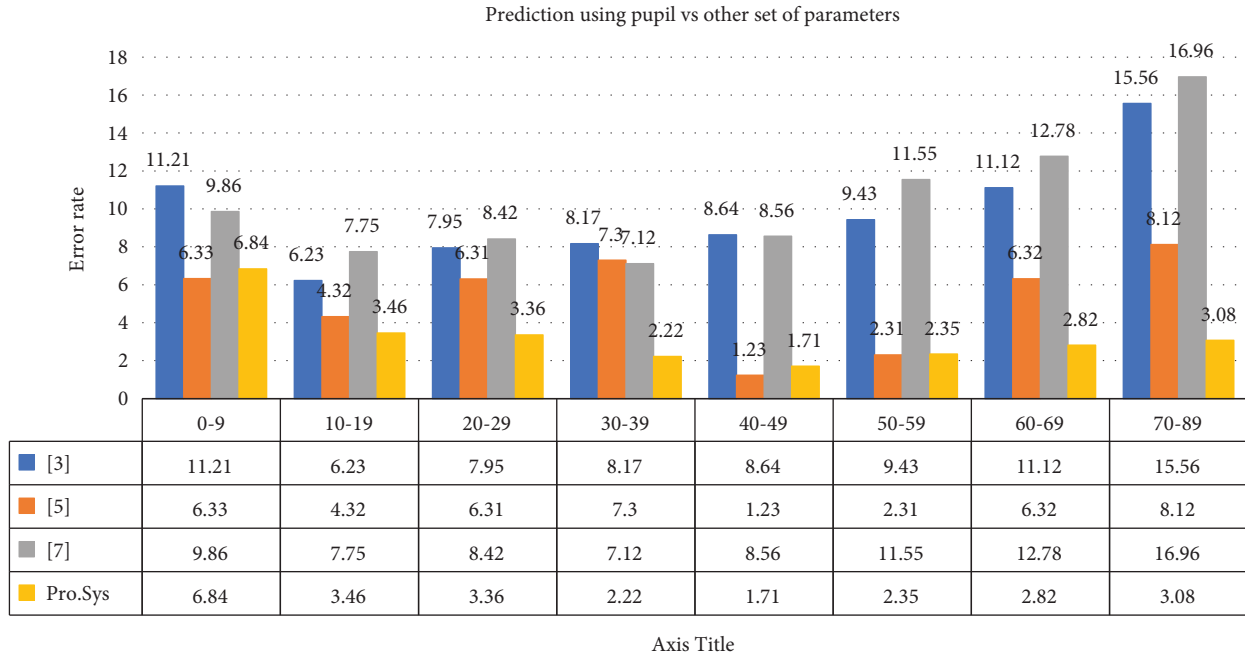


FIGURE 29: Comparison of an error rate of age prediction using Pupil with another set of parameters.

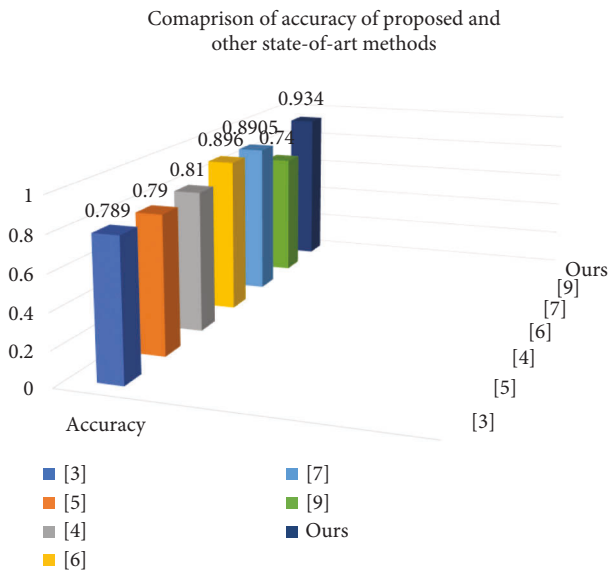


FIGURE 30: Accuracy comparison graph.

6.1. *Comparison Graphs of Parameters.* As stated in the previous section, the mean error and the accuracy were calculated for the proposed system and compared with other state-of-the-art methods to evaluate the performance of the proposed system. This section shows the comparison graph of the mean error rate of the proposed system against the state-of-the-art methods. When compared with the error rates of various age groups, as shown in Figure 28, the youngest age groups have an error rate compared to the old age subjects.

We compare the state-of-the-art methods of various parameters such as the Subject’s wrinkles, face, and

molar data. We gave a comparison chart of the error rate for all these parameters in Figure 29. There is a fair set of distinguishing error rates using the proposed methodology. Compared with the other existing methods, the proposed system shows better accuracy. Figure 30 compares the existing methods and the proposed method’s accuracy.

7. Conclusion

The proposed method is designed given the anonymous data we receive in forensic departments and criminal analysis. There were several methods to retrieve the age of the person. However, this proposed methodology is unique in predicting the Pupil’s age. The method has proven smooth implementation with an approximate accuracy of 95.6% on the random Subject. The proposed algorithm is compared with pretrained models such as ResNet and VGGNet to perform the efficiency check. The optimizers such as Gradient Descent, RMS Prop, and ADAM were used to check the Accuracy, Precision, *F*-score, and Specificity. We predicted the hyperparameter values of the proposed system using Adam as, Precision- 93.0%, Specificity- 96.4%, *F*-score 94.7%, and Accuracy- 96.0%, which were noted as higher compared to the other optimization techniques. ADAM was selected to perform the segmentation as the proposed system needs the accurate segmentation of pupils to predict the person’s age. Benchmark datasets such as CASIA, MMU, UBIRIS, and a few random images were collected to exhibit the system’s efficiency. The random dataset is chosen from Internet sources, where the image clarity and noise are almost near zero. Hence the resultant experimentation gave us a fruitful result. The experimentation succeeded while executing the proposed methodology on a random dataset,

which increased our confidence level in experimenting on live video. The classification accuracy for CASIA, UBIRIS, MMU, and random datasets was 89.6, 95.3, 93.1, 95.62, and 93.40, respectively. The proposed system has more than 20% accuracy compared to the other state-of-the-art methods. We want to extend the study to the noisiest datasets or else an excuse for training the system to generate the most accurate results. Age estimation based on biometric traits is attracting many researchers. One of the safe-to-use biometric traits is an eye. Eye not only authenticates a person but also hides much information needed in most applications. This application can be used in future security verification scenarios such as passport verification, border applications, age-restricted goods dispatch applications, etc.

Data Availability

The data used to support the findings of this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] K. Zhang, C. Gao, L. Guo et al., "Age group and gender estimation in the wild with deep RoR architecture," *IEEE Access*, vol. 5, Article ID 22492, 2017.
- [2] M. Erbilek, M. Fairhurst, and M. C. Da Costa Abreu, "Age prediction from iris biometrics," in *Proceedings of the 5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013)*, pp. 1–5, London, UK, December 2013.
- [3] S. Kim, Y.-H. Lee, Y.-K. Noh, F. C. Park, and Q. S. Auh, "Age-group determination of living individuals using first molar images based on artificial intelligence," *Scientific Reports*, vol. 11, no. 1, pp. 1073–1111, 2021.
- [4] F. Gao and H. Ai, "Face age classification on consumer images with Gabor feature and fuzzy LDA method," in *Proceedings of the International Conference on Biometrics*, Springer, Alghero, Italy, June 2009.
- [5] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 112–119, Miami, FL, USA, June 2009.
- [6] O. Agbo-Ajala and S. Viriri, "Deeply learned classifiers for age and gender predictions of unfiltered faces," *The Scientific World Journal*, vol. 2020, pp. 1–12, 2020.
- [7] S. Huang, N. Haiminen, A. P. Carrieri et al., "Human skin, oral, and gut microbiomes predict chronological age," *mSystems*, vol. 5, no. 1, pp. 006300-19–e719, 2020.
- [8] V. Sheoran, S. Joshi, and T. R. Bhayani, "Age and gender prediction using deep CNNs and transfer learning," in *Proceedings of the International Conference on Computer Vision and Image Processing*, pp. 293–304, Prayagraj, India, December 2020.
- [9] N. Sharma, R. Sharma, and N. Jindal, "Prediction of face age progression with generative adversarial networks," *Multimedia Tools and Applications*, vol. 80, no. 25, Article ID 33911, 2021.
- [10] M. M. Lavezzo, S. A. Schellini, C. R. Padovani, F. E. Hirai, and F. E. Hirai, "Evaluation of pupillary diameter in preschool children," *Neuro-Ophthalmology*, vol. 33, no. 6, pp. 313–317, 2009.
- [11] J. Zeng, H. Ling, L. Jan Latecki, S. Fitzhugh, and G. Guo, "Analysis of Facial Images across Age Progression by Humans," *International Scholarly Research Notices*, vol. 2012, Article ID 505974, 7 pages, 2012.
- [12] B. Winn, D. Whitaker, D. B. Elliott, and N. J. Phillips, "Factors affecting light-adapted pupil size in normal human subjects," *Investigative Ophthalmology & Visual Science*, vol. 35, no. 3, pp. 1132–1137, 1994.
- [13] H. Husniye, H. Erdol, and A. Turk, "The effects of age on pupil diameter at different light amplitudes," *Beyoglu Eye J*, vol. 10, pp. 31–37, 2018.
- [14] J. N. Sari, A. N. Hanung, E. N. Lukito, P. InsapSantosa, and R. Ferdiana, "A study on algorithms of pupil diameter measurement," in *Proceedings of the 2nd International Conference on Science and Technology-Computer (ICST)*, pp. 188–193, Yogyakarta, Indonesia, October 2016.
- [15] S. Gowroju and S. Kumar, "Robust deep learning technique: U-Net architecture for pupil segmentation," in *Proceedings of the 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 0609–0613, Vancouver, BC, Canada, October 2020.
- [16] A. Swathi and S. Kumar, "A Smart Application to Detect Pupil for the Small Dataset with Low Illumination," *Innovations in Systems and Software Engineering*, vol. 17, pp. 1–15, 2021.
- [17] A. Swathi and S. Kumar, "Review on pupil segmentation using CNN-region of interest," *Intelligent Communication and Automation Systems*, vol. 1, pp. 157–168, 2021.
- [18] S. Gowroju and S. Kumar, "Robust pupil segmentation using UNET and morphological image processing," in *Proceedings of the 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pp. 105–109, Cairo, Egypt, October 2021.
- [19] L. L. Lobato-Rincón, M. D C. Cabanillas-Campos, C. Bonnin-Arias, E. Chamorro-Gutiérrez, A. Murciano-Cespedosa, and C. Sãñchez-Ramos Roda, "Pupillary behavior in relation to wavelength and age," *Frontiers in Human Neuroscience*, vol. 8, no. 2, p. 221, 2014.
- [20] N. R. Misra, S. Kumar, and A. Jain, "A review on E-waste: fostering the need for green electronics," in *Proceedings of the 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 1032–1036, Noida, India, November 2021.
- [21] A. Jain and A. Kumar, "Desmogging of still smoggy images using a novel channel prior," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 1161–1177, 2021.
- [22] S. Kumar, A. Jain, A. Kumar Agarwal, S. Rani, and A. Ghimire, "Object-based image retrieval using the U-Net-Based neural network," *Computational Intelligence and Neuroscience*, vol. 2021, no. 2, 14 pages, Article ID 4395646, 2021.
- [23] A. K. Agarwal and A. Jain, "Synthesis of 2D and 3D NoC mesh router architecture in HDL environment," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 11, no. 4, pp. 2573–2581, 2019.