

Linear Regression Based Machine Learning Model for Cataract Disease Prediction

B. Premalatha¹, and Chandrasekharan Nataraj^{2*}

¹Assistant Professor, Department of Electronics and Communication Engineering, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

²Senior Lecturer, School of Engineering, Asia Pacific University of Technology & Innovation (APU), Technology Park Malaysia 57000, Kuala Lumpur, Malaysia

Abstract

Blindness and visual impairment have become major health problems today, with one major source of impairment being the formation of cataract disease. Cataract disease affects around 20 million individuals worldwide, with three out of every four being above the age of 60. A cataract is a clouding of the lens that impairs vision and can lead to blindness. Accurate and convenient detection of cataracts is essential for improving this condition. This paper presents a model based on linear regression to predict the presence of cataracts from images of the human eye. Experiments were conducted to validate the model. Various image processing techniques have also been used to perform image preprocessing and contrast enhancement, with patients notified about their eye health via SMS.

Keywords: Machine Learning Model, Linear Regression, Image processing, Feature Extraction, Cataract disease prediction.

Received on 11 02 2024, accepted on 28 05 2024, published on 24 06 2024

Copyright © 2024 B. Premalatha *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.10.6422

*Corresponding author. Email: chandrasekharan@apu.edu.my

1. Introduction

Visually oriented humans rely on their eyes to observe their surroundings. Today, however, blindness and visual impairment have become major health issues for humans, and the problem is significant worldwide. One of the most severe eye problems is cataract disease, which causes poor vision, an increased risk of falling, and also depression. Although uncommon, childhood cataracts leading to blindness and severe visual impairment are increasingly becoming a concern in health care. The primary strategic objective is early detection of the cataract disease, enabling self-administration of cataract tests. This model aims to prevent the requirement for remote residents to visit an ophthalmologist for cataract screening. The proposed model uses machine learning to predict cataract disease based on retinal fundus images of the human eye, which are analysed

using several image processing techniques as well as the machine learning model. The patient's cell phone receives an SMS containing the outcome's expected outcome.

2. Literature Survey

A review of existing literature enhances understanding of previously available techniques and provides insights for the efficient development of the system. In their 2019 study entitled "Cataract Detection using Digital Image Processing", Ishitaa Jindal et al. [3] proposed a cataract detection system based on digital image processing. This study discusses two algorithms: an area-based method for assessing cataract severity and an automated feature-based method for cataract identification and categorization. Two algorithms employ image processing techniques such as Hough transform and contour detection, concluding that a

feature-based technique which utilises histogram processing is more efficient.

Shruthi Bhat et al. [1] proposed a model in their work named "Cataract eye prediction using machine learning", based on convolutional neural networks (CNNs) using normal lens image input. The training model is built using a five-layer convolutional neural network that classifies the datasets as cataract and non-cataract by splitting the images into training and testing sets in order to make predictions.

An article on grading image modalities using machine learning [2] offered a thorough overview of the most recent developments in machine learning techniques, as well as their benefits and drawbacks. The survey gives a good overview of the different types of ophthalmic images that were used, as well as the methods of categorization and grading, and machine learning techniques that were employed. Thus, the survey concludes that DSLR or retinal photos combined with simple ML approaches can offer a quicker and more affordable method of detecting cataracts.

In their article titled "Automatic Cataract Classification Using Deep Neural Network with Discrete State Transition" [4], Yue Zhou et al. [4] developed a method based on neural networks with a discrete state transition. For network training using uniform discrete weights and activation, a DST technique was suggested. The paper concludes by saying that combining visible structures with HAAR features - and a few DST features - will help in identifying cataracts even in complicated images, as they could provide an accuracy of 92.85%. Based on the survey, it has been observed that there is a huge need for a time- and cost-effective model for cataract detection, which will help reduce the possibility of global blindness problems. By using proper image pre-processing methods along with the extraction of features, mainly mean, standard deviation, and light-weighted ML classifiers, better performance accuracy can be achieved.

In "Automatic Cataract Diagnosis by Image-Based Interpretability" [5], Jianqiang Li et al. built a model that offers a strong argument for its usage by localising the areas in the image that exhibit cataracts. The model makes use of a convolutional neural network as part of a deep learning method, which inputs a retinal fundus image and outputs it as a heatmap. In order to increase accuracy, the model was trained using two different techniques, the Adam and Frozing approaches. As a result, the ResNet flexible neural network was used in the model, offering 97.2% accuracy and overcoming problems with illness attribute localization.

Reddy Pavan et al. [6], in a paper named "Automatic Cataract Detection of Optical Image using Histogram of Gradient", proposed an automated detection mainly based on the Histogram of Gradient method. The proposed system was able to achieve better accuracy through segmentation methods, feature extraction and SVM classification to determine the severity of the cataract found in the eye.

Vinayak Holenavar et al. [7], in their survey titled "Instant Detection of Cataracts", discussed various types of cataracts, their region of occurrence and the effectiveness of current methodologies to detect them. This paper briefly explains age-related cataracts, edge cataracts and diabetic

cataracts, their causes, and their effects. The paper concluded that a better methodology would be based on features like histograms and mean intensities extracted from the images and compared with datasets that would identify the presence of cataracts.

In their study titled "Recent Approaches for Automatic Cataract Detection Analysis Using Image Processing," B. Ramesh Kumar et al. [8] explored the clinical evaluation image processing approaches as well as their benefits and drawbacks. Since retinal image techniques allowed for the quick and inexpensive examination of numerous images, they were the focus of the survey.

Based on the survey, it has been observed that there is a huge need for a time- and cost-effective model for cataract detection, which will help reduce the chance of blindness problems all over the world. By using proper image pre-processing methods along with the extraction of features, mainly mean, standard deviation, and light-weighted ML classifiers, better performance accuracy can be achieved.

3. Proposed Cataract Disease Prediction System

3.1. Linear Regression-Based Prediction Model

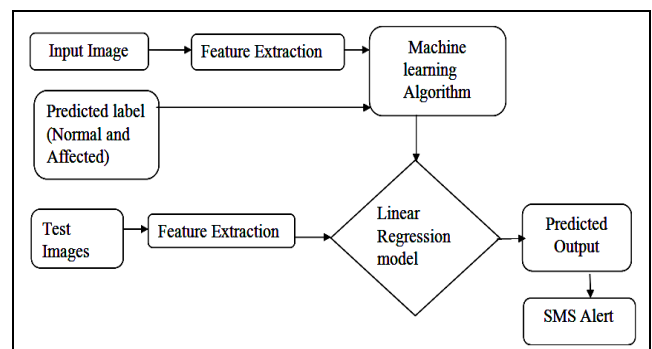


Figure 1. Linear Regression based Cataract disease prediction system

One of the most common eye conditions that impairs vision are cataracts. Early identification of cataracts is seen as a key strategy in the prevention of vision loss. Using image processing and machine learning models, a cataract state can be differentiated from a normal condition inside the pupil because it has a yellowish appearance. Figure 1 depicts the suggested model for predicting cataract disease.

The block diagram of a system for predicting cataract disease based on linear regression is shown in Figure. 1. The retinal fundus image used as input is from the Kaggle website. Pre-processing is done on the input image, and metrics like mean and standard deviation are computed. A linear regression model is designed using the extracted feature values. Once the system is trained for the given dataset, a new image is given as input, and predictions are made during the testing phase. After the prediction of the

presence of cataracts, an SMS alert will be sent to the respective patient's mobile number.

3.2. Input Retinal Images

The model for cataract prediction receives the retinal fundus image as input. Fundus retinal imaging, which captures a digital image of the back of the eye, can be used to determine a person's essential health state. The macular, optic disc and blood vessels make up the retina's major structural components. Figure 2 depicts a fundus retinal image. The optic disc, which is the enormous circular portion of the retina, also serves as an indicator of brain health, due to its close relation with brain nerves. The macular, which is the core of the eye and is yellow, contains abundant lutein. At the optic disc, blood vessels are linked and expand into arteries and veins. Veins are darker and stronger, while arteries are thinner and lighter than veins.

3.3. Machine Learning and Feature Extraction

The discipline of employing algorithms to process stored data, learn from it, and then anticipate what will happen based on gathered data is referred to as machine learning. The three machine learning algorithms accessible are supervised, unsupervised, and reinforcement learning. In supervised learning, also known as deterministic learning, a label is present for a certain dataset (training set) and based on the labels, the forecasts must be finalised. Other than labels, the missing result can be utilised once again in training. Finding the implicit associations in each unlabelled dataset is a challenging task in unsupervised learning. Feedback is supplied for each predictive action in reinforcement learning, which falls between supervised and unsupervised learning, but there is no specific classification.

3.4. Linear Regression Model

Linear regression, a supervised learning-based machine learning algorithm, performs a regression task to predict differences between retinal pictures with and without cataracts. The hypothesis function is trained to fit the gathered data effectively. The linear regression technique determines the linear relationship between the input and output variables, x and y . Equation (1) is the hypothesis function, used as a prediction model:

$$y = a+bx \quad (1)$$

Here, y denotes the output of the training data represented by x . During training, linear regression identifies the best line (represented by a and b) to predict the output y for a given value of the input x . By identifying the best a and b values, this model produces the best regression fit line. The retinal image used by the model to make predictions after training is known as the test image. By examining where the image feature values fall on the regression line, it is possible to determine whether the test image is normal or exhibits

signs of cataracts. After completion of the prediction, the patient receives an SMS via the Twilio module within the Spyder environment. The message includes information about their condition and appointment time. Multiple linear regression is a model with more than one independent variable with a single dependent variable. Linear regression is utilized in fields such as business, marketing, healthcare and medicine to create forecasts and make informed decisions.

3.5. Image Processing Operations

An overview of the picture pre-processing methods employed in the model for this paper is provided in this section. Pre-processing is crucial in biomedical pictures to ensure that segmentation and feature extraction algorithms produce accurate predictions. In the pre-processing stage of image processing, undesirable distortions are suppressed while the image quality is improved and enhanced for subsequent image processing processes.

3.5.1. Image Re-Sizing: In general, the term "image re-sizing" refers to scaling of images. By lowering the pixel count, it can either scale down or scale up the images to fit the need while also lowering their complexity. The image will be resized to a specific width and/or height using the `imutils` image resize module. It ensures that the aspect ratio is preserved and that the computer does not need to explicitly calculate the image's dimensions. To keep all the input images in the same dimension, image resizing is implemented in the image process flow. All the input photos in this model are shrunk to the same width of 500 pixels.

3.5.2. Grayscale Conversion: A white film that has formed inside the lens is often how cataract images are recognised. Grayscale images only feature a single layer with pixel values of 0-255, but colour images have three levels, each having pixel values of 0-255. Additionally, grayscale image processing is simpler than colour image processing. By converting the RGB values (24 bits) into grayscale (8 bits), a colour image can be turned into a grayscale image. The key idea behind this conversion is to convert the weighted sum of RGB pixel values to a single value, which denotes different shades of grey, including black and white. The OpenCV colour conversion module's `cv2.cvtColor_BGR2GRAY` flag is used to carry out this conversion. In Figure 2, the retinal fundus image has been converted to grayscale. It is simpler to perform additional processing on this image to find cataracts.

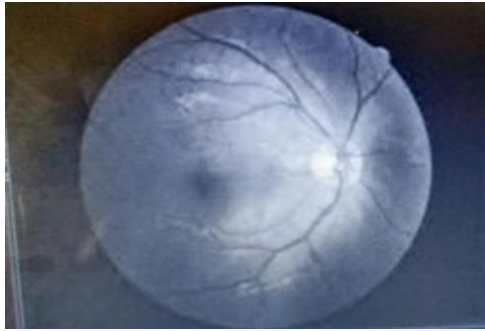


Figure 2. Grayscale converted image

3.5.3. Filtering: In Image processing, filters are used to enhance, smooth, and/or identify image edges. Any type of filtering includes applying an algorithm to the values of the pixels near the matching input pixel to determine the value of any given pixel. The drawback of these filters is that they also cause the edges to be smoothed. The photos should be noise-free and have sharpened edges for cataract diagnosis. In this study, the edge-preserving and noise-removal properties of the bilateral filter are chosen. Equation (2) can be used to formulate a bilateral filter.

$$BF[I]_p = \frac{1}{W_p} \sum_{q \in \Omega} G_{\sigma_s}(\|p - q\|) \times G_{\sigma_r}(\|I_p - I_q\|) I_q \quad (2)$$

where $(1/W_p)$ = normalization factor,
 $G_{\sigma_s}(\|p - q\|)$ is the spatial weight,
 $G_{\sigma_r}(\|I_p - I_q\|)$ is the range weight
 I represents the input image I ,
 p and q are spatial coordinates.

These conditions guarantee that only images with intensity values comparable to the centre pixel are considered and maintain crisp intensity changes.



Figure 3. Retinal Image after applying Bilateral Filter

This model employs the OpenCV bilateral filtering module. To produce edge-preserved and noise-free images, the argument values for d , sigma colour (colour space), and sigma colour (coordinate space) are selected as 15, 15, and 15, respectively. Figure 3 shows the grayscale image after applying a bilateral filter having the argument values (15, 15, 15).

3.5.4. Histogram Equalization: In image processing, histogram equalisation is used to boost the contrast. Contrast

enhancement is needed in image processing to fit the information within the grey-level image's limited dynamic range. It aids in enhancing the image's interpretability and consequently heightens the apparent contrast between the image's features. The histogram equalisation method was chosen because it broadens the histogram's dynamic range, which improves contrast. The intensity values of the pixels in the input pixels are assigned in a way that results in a uniform histogram to increase contrast. There are other ways to accomplish it, and this approach involves the calculation of the cumulative distributive function (cdf).

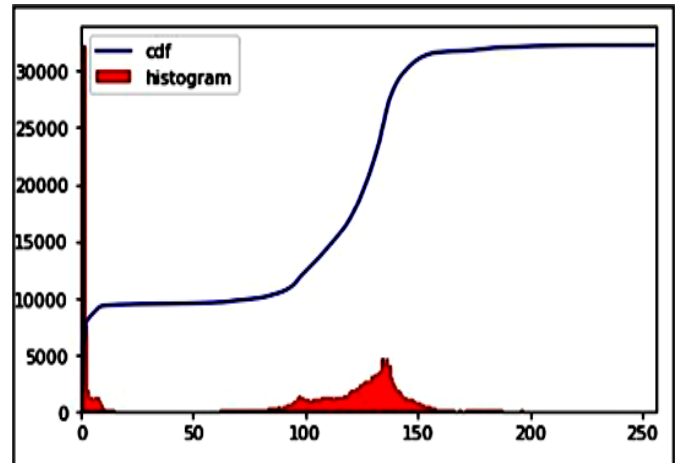


Figure 4. Histogram of the grayscale image with cdf evaluation

All the pixel values are normalised to fall inside the $[0, 1]$ range. A transformation is used to create a linear, flat histogram in the output image after computing the CDF of the input pixels. The process can be summed up as mapping the brighter input pixels to the entire output pixel region. The histogram of the grayscale image with the CDF evaluation is shown in Figure 4. Thus, using the OpenCV, Imutils, and Numpy libraries in Python, all input retinal images are resized to a uniform width of 500 pixels, transformed to grayscale to avoid complexity, filtered with a bilateral filter to maintain the edges, and boosted in contrast to discern features.

3.5.5. Feature Extraction: Inside the pupil, a yellowish tint is what a cataract looks like. The feature extraction method can aid to distinguish this disease from a normal condition that does not have a yellowish colour inside the pupil. This approach is based on analysing an image's texture properties. Mean and standard deviation are the textural features that were employed for the analysis.

The next section gives information about features used for analysing the images of both cataracts and normal eyes with a mathematical contribution.

4. Mathematical Modelling

4.1. Mean Intensity

Since the whitish hue of cataract eyes comes from the lens area, it can be inferred that they are more intense than normal eyes. Equation (2) provides the formula to calculate mean intensity (m).

$$m = \sum_{i=0}^{L-1} \left(\frac{I(i)}{N} \right) \quad (2)$$

Here, " m " stands for average intensity, " I " for maximum intensity, " N " for the total number of pixels in the image, and " L " for the total number of intensity levels. Both smooth and rough textures are visible to the human eye. Clouding suggests a rough texture, which in turn implies the presence of a cataract. The severity of the cloud cover is determined by its density.

4.2. Standard Deviation

In image processing, pixel values that have a very low standard deviation tend to be extremely close to the average value, whereas those that have a high standard deviation are dispersed throughout a wide range of values. Equation (3) provides the standard deviation calculation formula.

$$S = \frac{\sqrt{\sum (x_i - \bar{x})^2}}{r \times c - 1} \quad (3)$$

In Equation (3), " x_i " stands for the individual pixel value of the input map that the filter is considering, " \bar{x} " for the mean of the pixel values that the filter is considering, " r " for the filter size (in rows), and " c " for the filter size (in columns). The suggested approach is strengthened by the standard deviation calculation, which also incorporates an expert's diagnostic opinion and aids in finding almost exact pixel values in comparison to nearby pixels.

4.3. Dataset Creation

The features are calculated for a collection of datasets consisting of both cataract and normal images. All the computed values were automatically displayed in an Excel workbook by passing values to the sheet. It is possible to retrieve the information from an Excel spreadsheet using XLRD syntax in a simple manner.

- Step 1: Reading an Excel file using Python
- Step 2: Give the location of the file.
- Step 3: Activate the workbook
- Step 4: Determining the number of rows
- Step 5: Counting the columns to extract

The dataset's means and standard deviations are displayed in Table 1 using an Excel workbook.

The average value is then calculated by dividing the entire intensity of an image by the total number of pixels that make up the image.

As a result, the standard deviation aids in the detection of almost pixel values as compared to the surrounding pixel values, and the mean value provides the overall measure of the intensity of an image to the total number of pixels contained in it. To forecast the presence of cataracts in the human eye, these traits are employed to distinguish the different textures in a picture.

Table 1. Sample values of extracted features

	A	B	C
1	1	24.28885	29.03797
2	2	48.12535	48.62825
3	3	74.2407	74.37228
4	4	68.47266	69.01276
5	5	49.63923	52.76683
6	6	38.80714	42.4884
7	7	38.69199	45.90615
8	8	46.24483	47.25967
9	9	52.71693	53.81801
10	10	54.62683	57.66995
11	11	58.07625	58.48505
12	12	73.62503	77.07642
13	13	66.9376	67.38964
14	14	61.18529	64.13589
15	15	51.69302	55.22113
16	16	29.1452	33.28932
17	17	48.15055	50.81827
18	18	37.30117	40.85685
19	19	34.76191	39.20502
20	20	49.5073	51.12773
21	21	50.1192	53.76223

5. Mathematical Contribution to Form a Linear Regression Model

One of the most used techniques for predictive analysis is linear regression. Depending on the two variables' connection, to determine the link between the two variables, linear predictor functions are ones from which the unknown model parameters are calculated from input data. These models are known as "linear models". A simple linear regression method is one involving only a single input variable (x). Different techniques are available to form or train the linear regression equation from the dataset. The most common is called the Ordinary Least Squares technique. Fitting a linear equation using one variable as an explanatory variable and the other as a dependent variable can reveal the link between the two variables. The main goal is to find the relationship between x (an independent variable) and y (a dependent variable).

$$y = \beta_0 + \beta_1 x + \epsilon \quad (4)$$

where y : dependent variable, x : independent variable, β_0 : intercept, β_1 : gradient, and ϵ : residual error.

Here, β_0 and β_1 are termed "regression coefficients". The unobservable error component ϵ accounts for the data failure by representing the difference between the true and observed realisations of y . The difference between the predicted value and the observed value of an individual can be given by the

residual term. They are independently and identically distributed with zero mean and variance and account for nature's variability. The deterministic component is in the form of a straight line, and it provides the predicted response.

5.1. Fitness of the Model

Machine learning algorithms are mostly used in applications where there is a need to maximize or minimize a function. The group of functions that are to be minimized are called loss functions, which are a measure of how well a prediction model can predict the expected outcome. The most common method to find the minimum point of a function is gradient descent. The loss functions are of two types: (i) classification loss and (ii) regression loss. Learning algorithms are mostly used in applications where there is a need to maximize or minimize a function. The group of functions that are to be minimized are called loss functions, which are a measure of how well a prediction model can predict the expected outcome. In this work, the regression loss function is used to check fitness.

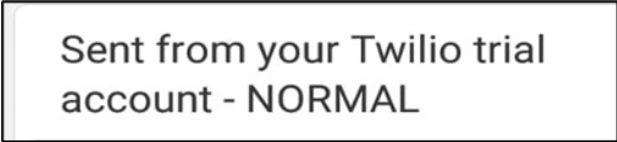
5.2. SMS Interface

Using training and testing datasets, the linear regression model is developed in this work. A new input image that must be predicted is fed into the model to obtain the output. This section gives an overview of the SMS module that is used in the model to send the predicted output to the patients.

Python allows sending SMS from an application with the help of Application Programming Interface (API). The Twilio module is used to send the SMS to the user. This model uses a REST API interface using HTTP's request and response. The client creates a Twilio account to send SMS to the users with a unique SM ID, which will be displayed in the Python application. The user receives the messages as 'cataract' or 'normal' and can take the necessary steps after the same.

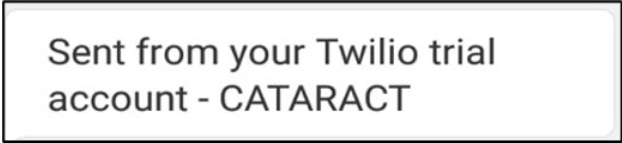
5.3. Creation of TWILIO Account

1. Install the Twilio module in Python and import the client from Twilio. rest
2. Get the account_sid and auth_token from the account.
3. Obtain the unique number.
4. Messages can be sent using the client. messages. create module.



Sent from your Twilio trial
account - NORMAL

Figure 5. Message Received-Normal



Sent from your Twilio trial
account - CATARACT

Figure 6.Message Received-Cataract

Figures 5 and 6 show the SMS that is received. Appropriate actions should be taken by the receiver after receiving the message. The prediction of cataracts is done using the trained linear regression model, and the predicted output is sent to the user via Twilio as an SMS.

6. Results and Discussion

A free open source Spyder IDE supported with Python platform was used to create the experiments. The test and train images are obtained from the Kaggle dataset. A folder containing 39 images is created, serving as a test and training dataset, consisting of 15 cataract images and 24 normal images. The Python modules used are xlswriter, CV2, imutils, NumPy, glob, and pandas. Matplotlib and Scikit-Learn are some of the libraries used. The images are read from the location mentioned in the code and pre-processed. The pre-processing phase involves image resizing, greyscale conversion, filtering, and contrast enhancement. With the help of the Glob module, the files or pathnames can be retrieved, and it's also a fats predictor.

A total of 39 images are considered for experimentation from the Kaggle dataset, and they have been processed sequentially. The imutils.resize() function is used to resize the width of the image and also tries to preserve the original aspect ratio of the image. In this work, the images are resized to have a maximum width of 500 pixels and converted to grayscale before the bilateral filter is applied. By converting an RGB image to a grayscale image, non-processed data can be removed; the huge amount of data in an RGB image may not be required for processing. Grayscale images are entirely sufficient for many tasks compared to harder-to-process colour images. A bilateral filter is non-linear in its operation and is used as an edge-preserving and noise-reducing smoothing filter for images. The intensity of each pixel has been replaced by the weighted average of the intensities of its neighbours. The experimentation details carried out on a total of six images are shown in Figures 6(a) to 6(e).



Figure 6(a-1). Original Image of cataract eye-1

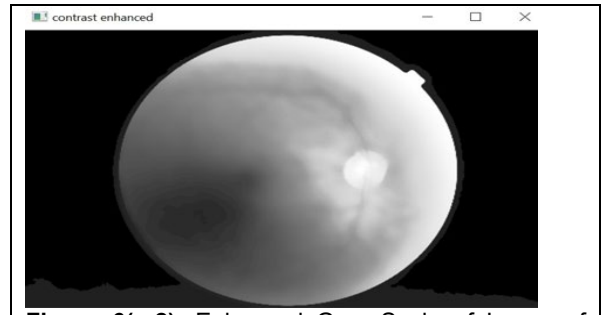


Figure 6(a-2). Enhanced Gray Scale of Image of Cataract eye-1

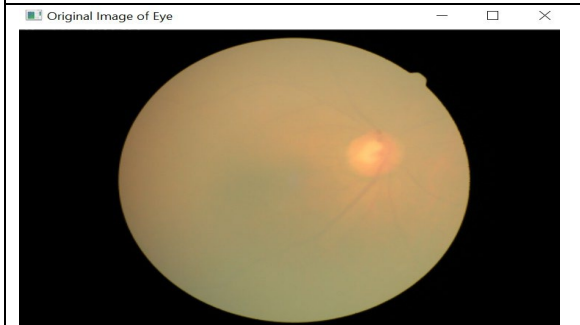


Figure 6(b-1).Original Image of Cataract eye-2

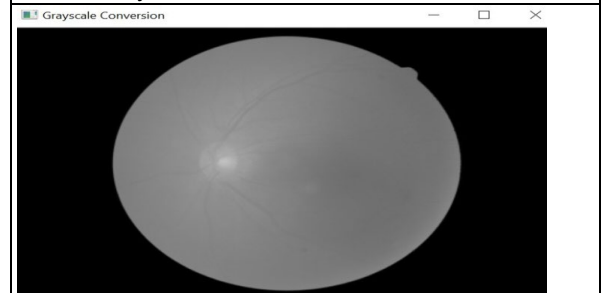


Figure 6(b-2). Enhanced Gray Scale of Image of Cataract eye-2



Figure 6(c-1).Original Image of Cataract eye-3

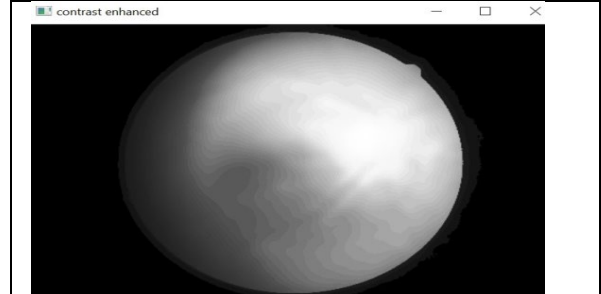


Figure 6(c-2). Enhanced Gray Scale of Image of Cataract eye-3



Figure 6(d-1).Original Image of Normal eye-1

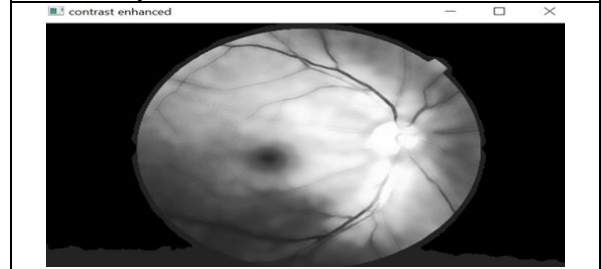


Figure 6(d-2). Enhanced Gray Scale of Image of Normal eye-1



Figure 6(e-1).Original Image of Normal eye-1

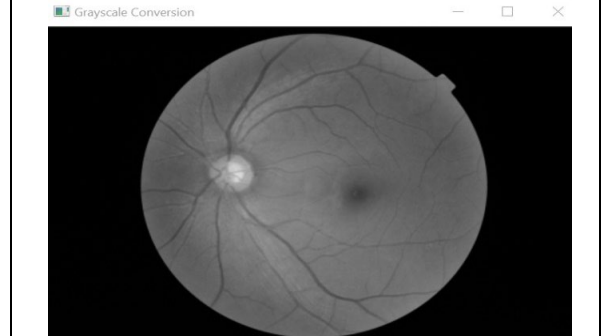


Figure 6(e-2). Enhanced Gray Scale of Image of Normal eye-2


```
[[1, 24.28884592145015, 29.037970143032023], [2, 48.12535347432024, 48.6282532857628],
[3, 74.24070090634441, 74.37227569718989], [4, 68.4726586102719, 69.01276151878294], [5,
49.639232628398794, 52.76682702432245], [6, 38.80714199395771, 42.48839531334528], [7,
38.69198791540786, 45.90614678717629], [8, 46.244827794561935, 47.25966931851175], [9,
52.716930513595166, 53.81800846192528], [10, 54.62683483483484, 57.669945670560445],
[11, 58.076253776435045, 58.48505165367111], [12, 73.62503303303303, 77.07642048622185],
[13, 66.93760120845921, 67.3896367824144], [14, 61.18529120129129, 64.13589338498782],
[15, 51.693015015015014, 55.22113046387663], [16, 29.145196374622355,
33.28931775090852], [17, 48.150549848942596, 50.81826543787759], [18,
37.301172205438064, 40.856848640113704], [19, 34.761909365558914, 39.20502467000207],
[20, 49.50729909365559, 51.12772526852831], [21, 50.1192012012012, 53.762227003118966],
[22, 50.7869009090909, 54.28704069086989], [23, 60.62564350453172, 62.33657509339458],
[24, 50.790259818731116, 55.39311296306735], [25, 51.80059818731118,
53.236266432672664], [26, 46.63989123867069, 48.12924117855191], [27,
57.899591591591594, 61.37304048465988], [28, 58.217975975975975, 61.49859931498545],
[29, 52.95615105740181, 53.75192321652939], [30, 54.24829003021148, 55.818060362528854],
[31, 35.2735166163142, 39.34597256696879], [32, 50.1212749244713, 53.60166627764595],
[33, 57.09090909090909, 61.33938499398758], [34, 63.40640840840841, 67.30067286395663],
[35, 48.33913595166163, 49.84180741368118], [36, 48.23613897280967, 51.01391732698177],
[37, 53.85154078549849, 55.30524678906509], [38, 36.00477341389728, 38.67263615118827],
[39, 47.05905740181269, 48.80641077529773]]
```

Figure 7. Array containing Mean and Standard Deviation of Input images

Figure 7 shows the output displaying the mean and standard deviation of 39 images. The prediction of cataract disease is done using the Linear Regression model. Based on the test size value, the data is divided into training and testing sets. Figures 8 and 9 depict the training and test sets when the test size is set to 0.5. To prevent overfitting and to appropriately evaluate the model, the data is divided into training and test sets. When the test size is set to 0.5, the data set is divided into two equal samples.

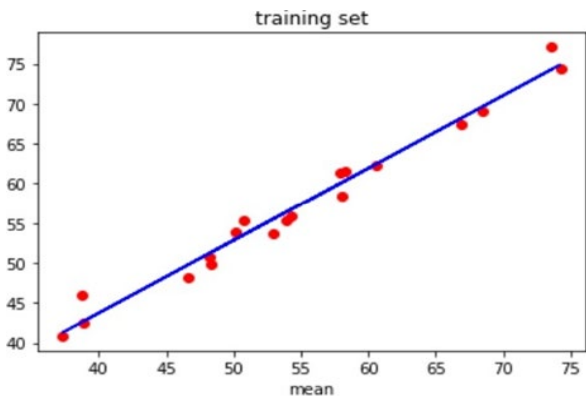


Figure 8. Training set when the test size is 0.5.

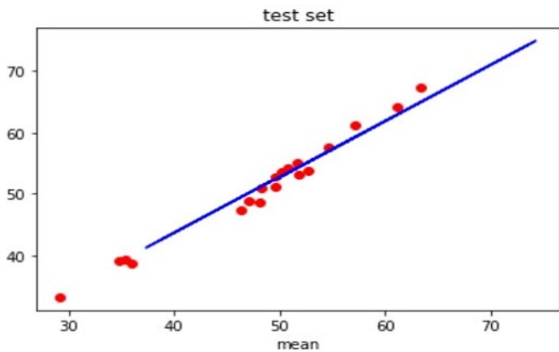


Figure 9. Test set when the test size is 0.5.

From the graph shown in Figures 10 and 11, it is observed that the points lying above the regression line are

normal images and the points lying below the line are cataract images.

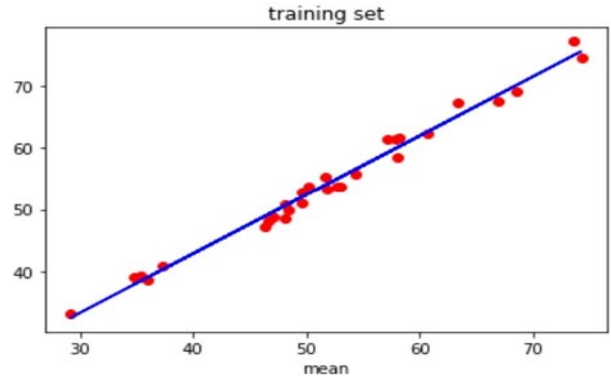


Figure 10. Training set when the test size is 0.2

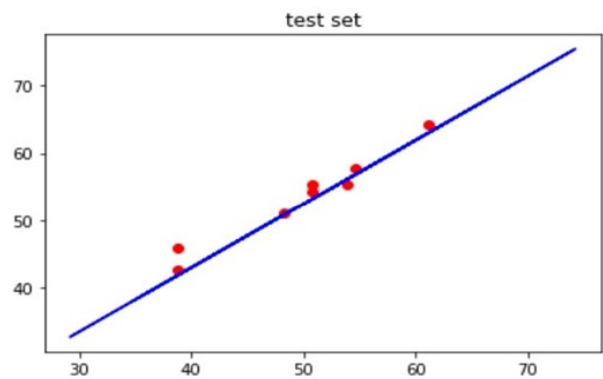


Figure 11. Test set when the test size is 0.2

Figure 12 shows the message received via the Twilio account, and "SMadbc827e3ba14b51b91ab7f97d004a4" in Figure 13 shows the acknowledgement obtained in Spyder.

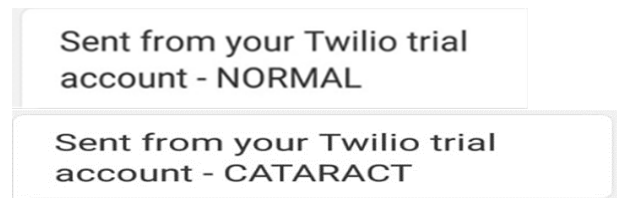


Figure 12. Message received-Normal and Cataract

```
[[1, 30.772996996996998, 33.468199616240845]]
[[1, 30.772996996996998, 33.468199616240845], [2, 59.377583081571,
60.92085514121874]]
predicted response [61.4799924] actual response [60.92085514]
SMadbc827e3ba14b51b91ab7f97d004a4

[[1, 59.377583081571, 60.92085514121874]]
[[1, 59.377583081571, 60.92085514121874], [2, 65.92293655589124,
67.33048649518845]]
predicted response [67.62296907] actual response [67.3304865]
SM3c9bd03a4e3942f8b78424215002236e
```

Figure 13. Acknowledgement obtained in Spyder for Normal and Cataract

The regressor intercept, regressor coefficient, regressor score, mean absolute error, mean squared error, and root mean squared error were calculated and shown in Figure 14. The extracted feature values are also displayed in Table 2.


```
regressor intercept 7.579957812246107
regressor coefficient [0.90629135]
Mean Absolute Error: 1.17997343147848
Mean Squared Error: 1.921165770709117
Root Mean Squared Error: 1.386061243491469
regressor score 0.9718940567950572
```

Figure 14. Calculated Regressor and Error-values

Table 2. Extracted Feature Values

Input Retinal Image	Mean	Standard deviation
	24.2888	29.0379
NORMAL Images	48.1253	48.6282
CATARACT Images	74.2407	74.3722
	47.0590	48.8064
	53.8515	53.3052
	48.236	51.0139

6. Conclusion

A cataract is the clouding of the lens in the eyes that occurs when protein inside the eyes clumps together, leading to a loss of crystalline lens transparency. There is a need to develop a model for the prediction of cataract disease in humans with better accuracy. This work will allow people who find it difficult to visit hospitals to check for the presence of cataract disease. When given an image as input, the process uses image processing and a linear regression-based machine learning model to predict the existence of cataract illness from the retinal fundus images of an eye. In the future, we can expand the dataset's sample size and provide an Internet of Things (IoT) interface for conveying messages to patients.

References

- [1] Shruthi Bhat, Som Mosalagi, Tejal Bhalerao, Pushpak Katkar, Rahul Pitale. Cataract Eye Prediction using Machine Learning. *International Journal of Computer Applications*. 2020; 176(35): 46-48.
- [2] Zhang XQ, Hu Y, Xiao ZJ, Fang JS, Higashita R, Liu J. Machine learning for cataract classification/grading on ophthalmic imaging modalities: A survey. *Machine Intelligence Research*. 2022;19(3):184-208.
- [3] I. Jindal, P. Gupta, and A. Goyal. Cataract Detection using Digital Image Processing. 2019 Global Conference for Advancement in Technology (GCAT), Bangalore, India. 2019; 1-4.
- [4] Y. Zhou, G. Li, and H. Li. Automatic Cataract Classification Using Deep Neural Network with Discrete State Transition. *IEEE Transactions on Medical Imaging*. 2020; 39(2):436-446.
- [5] Jianqiang Li, Xi Xu, Yu Guan, Azhar Imran, Bo Liu, Li Zhang, Ji-jiang Yang, Qing Wang, Liyang Xie. Automatic Cataract Diagnosis by Image-Based Interpretability. *IEEE International Conference on Systems, Man and Cybernetics*. 2018.
- [6] T. Reddy Pavan, Dr.A.Deepak. Automatic Cataract Detection of Optical Image using Histogram of Gradient. *International Journal of Engineering Research and Technology (IJERT)*, 2018; 7(6):383-385.
- [7] Vinayak Holennavar, Dr.Priyatam Kumar. Instant Detection of Cataracts. *International Journal of Latest Trends in Engineering and Technology*. 2018; 9(4):38-44.
- [8] B. Ramesh Kumar, Shimna M.P. Recent Approaches for Automatic Cataract Detection Analysis Using Image Processing. *Journal of Network Communications and Emerging Technologies (JNCET)*. 2017; 7(10).