ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2023.0420

AN ENHANCED ENSEMBLE HYBRID DEEP LEARNING ALGORITHM FOR IMPROVING THE ACCURACY IN IRIS SEGMENTATION

M. Sathiya[1], K. Karunambiga[2], G. Sai Chaitanya Kumar[3] and S. Chandra Sekaran[4]
[1]*Department of Information Technology, Karpagam Institute of Technology, India*
[2]*Department of Computer Science and Engineering, Karpagam Institute of Technology, India*
[3]*DVR and Dr. HS MIC College of Technology, India*
[4]*Department of Computer Science and Engineering, PSV College of Engineering and Technology, India*

Abstract
In recent years, there has been a meteoric rise in the application of deep neural networks for the purpose of iris segmentation. This can be attributed to the extraordinary capacity for learning possessed by the convolution kernels that are utilised by CNNs. Conventional methods have several drawbacks, some of which can be partially compensated for by using CNN-based algorithms, which increase the segmentation precision. On the other hand, the CNN-based iris segmentation approaches that are currently in use typically require a more complex network, which results in an increase in the number of parameters. This is essential to realise a higher degree of precision in the results. CNN-based techniques are effective, they can only be used for a specific application. This makes them inappropriate for general iris segmentation goals, even though they are effective.

Keywords:
Ensemble Model, Deep Learning, Iris Segmentation

1. INTRODUCTION

When developing an iris identification system, the first stage in the preprocessing phase must always be the segmentation of the iris [1]. If the segmentation of the iris is done incorrectly, it is possible for identity-related information to be erased as well as for new interferences, such as eyebrows and eyelashes, to be introduced. Both factors can contribute to a reduction in the precision of iris identification. Iris segmentation has several important applications, not the least of which is in the field of medicine, in addition to its use in devices that perform iris identification [2].

Before an effective computer-assisted ocular disease diagnosis can be made, the pre-processing pipeline must first be finished, which includes an important first stage known as accurate iris segmentation. This stage must be finished before the next stage in the pipeline can begin. The ability to divide the iris is clinically significant because it facilitates in the early discovery of neovascular glaucoma. This is one of the most common causes of blindness worldwide. The capacity to divide the iris is essential for a few reasons, but this is one of the most significant ones. As a direct consequence of these issues, several individuals have begun offering solutions in the form of suggestions for improved methods of iris segmentation. There are still a great many obstacles to overcome in this branch of research [3].

Iris recognition devices require input from a person, they are vulnerable to a wide range of restrictions on account of this requirement. Iris photographs, on the other hand, taken in settings with a less stringent degree of control are more likely to have an excessive amount of ambient noise in the background. This noise may be brought on by several different things, such as motion blur, occlusions brought on by the eyelids or eyelashes, occlusions brought on by spectacles, position offset, and other similar things.

Taking an image of an iris can be done with either visible light or near infrared photography, which are the two most prevalent approaches. The other technique that is frequently used is called near-infrared. The iris is responsible for producing an image, the clarity of which can be significantly influenced by factors such as the lighting and the background [4].

Eye segmentation techniques that are based on manual labour and convolutional neural networks respectively are the two primary categories of techniques that fall under the category of earlier methods. The Hough transform is the fundamental building block upon which most of the diverse segmentation approaches that are presently in use are constructed. The only circumstances in which these conventional techniques can produce segmentation results that are up to standard are those that are ideal. Conventional methods of iris segmentation will have a much lesser degree of accuracy if there is extraneous noise present in the image of the iris that is being segmented [5].

2. LITERATURE SURVEY

The steps that are involved in the conventional method of iris recognition are as follows: taking a image of a person iris, putting that image through a series of mathematical operations to generate a feature vector, using that vector to generate an iris code, and finally storing that code as a template. This method is used to identify individuals based on their eyes. During the authentication process, these iris templates are compared to a freshly created template of the same iris by making use of distance measurements. This comparison takes place between two templates of the same iris. Depending on the format in which the textured data is presented, the two primary techniques that have emerged for extracting iris features are the use of real-valued feature vectors and binary iris codes. The length of real-valued feature vectors is longer than that of binary retinal codes. The first device that was able to successfully recognise iris patterns was built with the help of the former method [6].

Daugman method for separating the iris from the rest of the eye depended on the assumption that the pupil is spherical, which is not always the case. This is because the iris is much more circular than the rest of the eye. There are a few distinct methods that can be utilised to dissect the iris from the remainder of the eye. In addition, real-valued feature vector-based techniques make use of transformations that are comparable to the binary iris code; however, the product of these transformations is real-valued vectors rather than binary codes [7].

Researchers have been utilising many different approaches to machine learning over the course of the past twenty years to successfully recognise individuals based on the patterns found in their retinas. The process of feature extraction, on the other hand, makes use of the local and (or) global techniques that were explained earlier. Following the process of constructing ocular feature vectors with Gabour wavelets, various kernels of support vector machines (SVMs) were evaluated on the same dataset. Using a gradient descent technique is one way to solve problems in artificial neural networks (ANN) In addition, the log-sigmoid function is implemented in the final layer of a three-layer ANN, which serves as the layer concluding layer [8].

The discrete cosine transforms (DCT), is applied in this layer to classify the feature vectors. The methods that were utilised for the purposes of classification using the Kohonen self-organizing map (SOM) and feature extraction using the Independent Component Analysis (ICA) [9].

To classification, a feature vector as well as a multilayer feed-forward network will be developed using the sclera-iris border contour points and the pupil-iris border contour points, respectively. In addition, the utilisation of SVM in conjunction with the hammering distance has the potential to improve the efficiency of the classification procedure. The Support Vector Machine (SVM) is utilised as the primary classification method. The Hamming distance is utilised as a secondary classification method if the SVM is unable to classify the data correctly. In recent years, convolutional neural networks, also known as CNNs, have been a game-changer in the field of computer vision since they permit self-feature construction. The reason for this is that CNNs can take in input from a variety of sources and generate their own output [10].

In addition, a hybrid CNN model was developed to determine the degree of similarity that exists between two periocular images that encompass the eye. This model was developed to determine the degree of similarity that exists between the two images. To achieve improved accuracy in image identification, using the pair-wise convolutional filter to analyse the degree of similarity between two iris images obtained from different sources is recommended. Despite this fact, the primary focus of these efforts was focused not on iris identification but rather on iris authentication. IrisConvNet is a multi-biometric algorithm that was based on deep learning and combines the iris of a person eyes to determine that person identity. It does this by combining the iris from both person eyes. On the other hand, no research was done into whether the iris detector could be used in conjunction with other sensors [11].

DeepIrisNet was successful in developing two distinct deep CNN models for iris recognition across a wide range of sensor types. However, a greater network depth may create issues such as disappearing gradients and weight saturation, neither of which are accounted for by these models since they do not consider the effects of increased network depth. The algorithms do not consider either of these concerns in any way. CNN models that have already been pre-trained with accurate recognition at each stratum of the network. MiCoReNet model for eye identification combines fundamental convolutional layers with residual layers to make the most of both the rapid convergence that can be achieved through convolution and the non-saturation that can be achieved through residual connections. This was done to take advantage of both the rapid convergence that can be achieved through convolution and the residual connections. On the other hand, individuals who have large pupils do not benefit all that much from iris image classification because it is not as accurate for them [12].

The literature review suggests that a deep-learning framework is required to manage cross-sensor iris recognition. This framework should have features such as optimal depth with more accurate layer configuration, the rapid convergence mechanism, non-saturation, and the ability to deal with large iris categories. The capacity of iris features constructed from a single viewpoint to differentiate between subjects with extremely large irises has been called into question by recent research. The situation has become even more complicated because of this development. Because the iris patterns of different people are completely unique to one another, it would be beneficial to make use of a collection of admirable characteristics to improve human discrimination.

3. ENSEMBLE LEARNING

Deep neural networks, of which convolutional neural networks are a subcategory, are what are used in the process of automatically extracting characteristics. The study decided to use CNN rather than the alternative because it has been shown to be more effective than traditional artificial neural networks in the process of extracting features from images.

CNNs can carry out a transformation on pixel matrices by making use of a collection of mathematical procedures that are collectively referred to as convolution. This collection of manipulations is what required to carry out the transformation.

CNNs make use of filters, which can assist in the process of extracting the spatial significance of various components present in an image, they are superior to other kinds of deep neural networks when it comes to their performance on image data. This is because filters can assist in the process of extracting the spatial significance of various components present in an image. This is since other varieties of deep neural networks do not make use of any filters.

3.1 FINE TUNING

It was determined that the binary cross-entropy would be the most efficient loss function to put into action, so that is the one that will be used. An illustration of the calculation for the binary cross entropy can be given as an example as follows:

$$Loss = -\frac{1}{N}\sum_{i=1}^{n} y_i \log\left(p\left(y_i\right)\right) + \left(1-y_i\right)\log\left(1-p\left(y_i\right)\right) \quad (1)$$

where

y_i - output label and

$p(y_i)$ - predicted probability.

The weights are updated.

$$w_t = w_{t-1} = \frac{\eta \hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}} \quad (2)$$

The first moment estimate, which is indicated by m_t, and the second moment estimate, which is denoted by v_t, for effective loss convergence.

$$\hat{m}_t = m_{t1} - \beta_{t1} \quad (3)$$

$$\hat{v}_t = v_{t1} - \beta_{t2} \quad (4)$$

where β_1 and β_2 - decay rates.

When performing an evaluation of each deep learning model, the learning rate, algorithm, and loss functions are held to the same standards throughout all the models.

3.1.1 InceptionV3:

InceptionV3 convolutional neural network architecture is a component that belongs to the Inception series of architectures for convolutional neural networks. When compared to earlier iterations, this edition has several improvements that can be found throughout. In the context of this discussion, the phrase these refers to intelligent factorised convolutions, label smoothing, and the utilisation of a supplementary classifier for the purpose of disseminating label information throughout the network. We retrained Google InceptionV3 architecture on our dataset by applying fine-tuning across all layers, replacing the top layers with one average pooling, four fully connected layers, and finally the sigmoid activation function in the output layer.

This was done after first replacing the top layers with four fully connected layers. This was accomplished by first removing the top layers, then adding one layer with an average pooling function, then four layers with complete connectivity, and finally adding the sigmoid activation function. As a result of this, we were able to position samples into one of two categories for the purposes of diagnostics: benign or malignant.

3.1.2 VGG19:

The VGG19 variant is precisely what one would anticipate based on its name; it is a convolutional neural network that has been learned in preparation and contains 19 layers. The one thousand distinct classes that are contained within the ImageNet collection serve as the foundation for the training of the VGG-19. This will function as the basis for the training moving forward. Before applying the sigmoid layer to obtain the classification result, we perform fine-tuning across all the layers that comprise the VGG-19 architecture as part of this research.

In addition, we swap the upper layers for four layers that are all interconnected with each other. To be more particular, the input is resized so that its measurements match those of the initial input that the VGG19 used. This is done so that the measurements can be compared. This is accomplished by ensuring that its breadth and height are identical to one another. Because of this, you can rest assured that it will function normally when connected to the network.

3.1.3 ResNet:

The dataset is retrained by making use of the ResNet ResNet50 architecture, and each layer is fine-tuned so that it can be utilised in various types of experiments. This enables the network to have a higher degree of precision. The higher layers are eliminated, and in their place comes a collection of four layers that are in full connection with one another and make use of RLU as their activation function. Following that, the sigmoid layer is applied so that assistance can be provided in the interpretation of the observations regarding the two diagnostic groups.

The network implements identity mapping to circumvent the problem of disappearing gradients and train much deeper networks. This enables the network to train more complex neural networks. The network can acquire more complex patterns because of this. This parameter-less identity mapping does nothing more than combine the findings from the layer below it with the information that is provided by the layer above it. When layers are stacked on top of one another, the outputs of the individual layers are combined by using the skip connections that are present between the layers.

The pretrained weights that are utilised by the ResNet50 technology have not been subjected to any sort of modification. The activation function has been changed to relu, the same as it was in the previously pretrained CNNs, and the top layer has been expanded to include four layers that are all completely connected to one another. This layer will make use of the sigmoid activation function, which is going to be the activation function for the very last output layer after this one.

3.1.4 DenseNet201:

There are 201 hidden layers dispersed throughout the construction of the convolutional neural network that is referred to as DenseNet201. The information that is needed for one layer in DenseNet201 is obtained from the layer that is located directly above it and is then used for the layer that is located directly below it. This design makes use of a straightforward connectivity structure, which serves to maximise the amount of data movement that can occur in both directions between layers. The problem of vanishing gradients can now be fixed as a direct result of this thanks to the construction of the game. Because each layer is connected to every other layer, the feature maps that are recorded in the layers that are on top of it can be used to retrieve the data from the layers that are below it. This is possible since every layer is connected to every other layer.

The entire structure has been broken up into several blocks that are all closely affiliated with one another to make it simpler to perform downsampling. This was done to make the process of downsampling easier to perform. Convolution and pooling are two processes that are carried out by the transition layers that are located between the dense chunks in the image. These transition layers are situated in between the layers that make up the image. We were able to build an additional four completely connected layers on top of this previously trained structure by utilising relu as the activation function. This allowed us to make the structure more complex. This structure serves as the foundation upon which everything else is built. When it comes to the process of categorization, we conduct it out by employing a sigmoid activation function on the very last layer of the network, which is the layer that contains all the network outputs.

3.1.5 InceptionResnetV2:

The model that is presently being presented to you is the result of combining two distinct algorithms, namely Inception and ResNet. Algorithms were used to create the model. When you have finished retraining the InceptionResNetV2 architecture with our data, we suggest that you conduct fine-tuning on each of the individual layers. The top layers of diagnostic images are stripped away and replaced with four new layers that are fully connected to one another across their entirety. This is done so that the images can be classified into two distinct categories. At the very end of the procedure, a sigmoid layer is added to finish things off.

After putting all these models through their paces on the ISIC dataset, which is comprised of skin cancer, we discovered that the VGG19 pretrained CNN performed the best out of the bunch. This dataset was used to test the model ability to identify the disease.

4. RESULTS

The databases have records of many different types of irises, including the type of iris shown in Fig.1, which can be found in the databases. The ND-Iris-0405 collection has a total of 64,980 iris samples, all of which were obtained by surgically removing them from the eyes of 356 different individuals. There are 712 different irises that have been produced, and the measurements of each one is 480x 640 pixels.

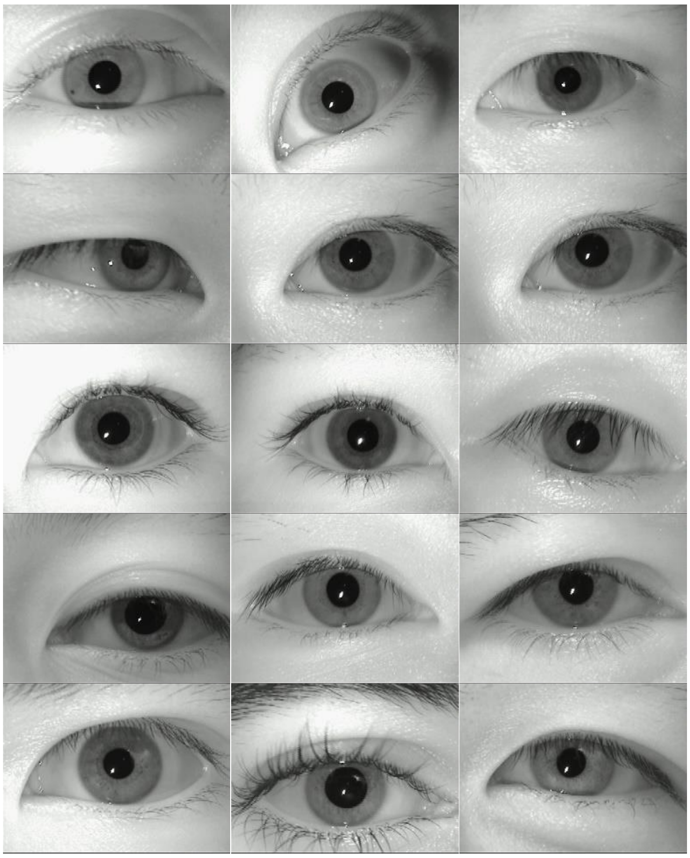

Fig.2. Image Samples

This section offers a graphical illustration of the experimental setting that was utilised in the evaluation of the proposed method that was presented earlier in the section. Iris samples were taken from both the left and right eyes of every individual that was included in both datasets; consequently, it is important that the samples be identified in a manner that is distinct from one another.

Although there is some variation in the distribution of iris images from one individual to the next, there is also a significant amount of variation between the left and right irises of any one person. The classification that is produced as a direct consequence of this is likely to be arbitrary because it is derived from such unprocessed information. After this step, each dataset is then subdivided into three subsets based on iris themes, with each subset comprising a progressively fewer number of images than the previous one.

Table.1. Simulation Parameters

Property	IITD	CASIA-v4	UBIRIS.v2
Image Size	320 × 240	320 × 280	400 × 300
Input Size	320× 224	256 × 256	384× 288
Number of images	1000	1524	2154
Training sets	1684	2054	1625
Validating sets	325	315	215
Testing sets	265	275	215

Table.2. Training

Database	Approach	Accuracy	Precision	Recall	MAE
CASIA-v4	CNN	0.8193	0.7767	0.9395	0.1492
	IrisConvNet	0.8000	0.6475	0.9257	0.2408
	DeepIrisNet	0.8213	0.9085	0.9293	0.0855
	MiCoReNet	0.9081	0.9122	0.9880	0.0683
	Proposed	0.9293	0.8818	0.9313	0.2408
IITD	CNN	0.9880	1.0014	1.0026	0.0122
	IrisConvNet	0.9739	0.9940	1.0023	0.0160
	DeepIrisNet	0.9907	1.0028	0.9995	0.0118
	MiCoReNet	0.9912	1.0030	1.0025	0.0108
	Proposed	0.9923	1.0035	0.9991	0.0115
UBIRIS.v2	CNN	0.9138	0.9665	0.4214	0.0526
	IrisConvNet	0.9260	0.3445	0.4324	0.0156
	DeepIrisNet	0.9353	0.9734	0.9988	0.0290
	MiCoReNet	0.9868	1.0008	1.0021	0.0127
	Proposed	0.9850	0.9999	0.9967	0.0121

Table.3. Testing

Database	Approach	Accuracy	Precision	Recall	MAE
CASIA-v4	CNN	0.8160	0.7737	0.9358	0.1486
	IrisConvNet	0.7968	0.6449	0.9220	0.2399
	DeepIrisNet	0.8181	0.9049	0.9257	0.0851
	MiCoReNet	0.9045	0.9086	0.9841	0.0681
	Proposed	0.9257	0.8783	0.9277	0.2399
IITD	CNN	0.9841	0.9974	0.9987	0.0121
	IrisConvNet	0.9700	0.9901	0.9984	0.0160
	DeepIrisNet	0.9868	0.9989	0.9955	0.0117
	MiCoReNet	0.9873	0.9991	0.9986	0.0107
	Proposed	0.9883	0.9996	0.9951	0.0114
UBIRIS.v2	CNN	0.9102	0.9627	0.4197	0.0524
	IrisConvNet	0.9223	0.3431	0.4307	0.0156
	DeepIrisNet	0.9316	0.9695	0.9948	0.0289
	MiCoReNet	0.9829	0.9968	0.9982	0.0126
	Proposed	0.9812	0.9959	0.9928	0.0120

Table.4. Validation

Database	Approach	Accuracy	Precision	Recall	MAE
CASIA-v4	CNN	0.7871	0.7463	0.9026	0.1434
	IrisConvNet	0.7686	0.6221	0.8894	0.2314
	DeepIrisNet	0.7891	0.8729	0.8929	0.0821
	MiCoReNet	0.8725	0.8764	0.9493	0.0656
	Proposed	0.8929	0.8472	0.8948	0.2314
IITD	CNN	0.9493	0.9621	0.9633	0.0117
	IrisConvNet	0.9357	0.9550	0.9630	0.0154
	DeepIrisNet	0.9519	0.9635	0.9603	0.0113
	MiCoReNet	0.9524	0.9637	0.9632	0.0103
	Proposed	0.9533	0.9642	0.9599	0.0110
UBIRIS.v2	CNN	0.8779	0.9286	0.4049	0.0505
	IrisConvNet	0.8897	0.3310	0.4154	0.0150
	DeepIrisNet	0.8986	0.9352	0.9596	0.0279
	MiCoReNet	0.9481	0.9615	0.9628	0.0122
	Proposed	0.9464	0.9607	0.9576	0.0116

Overfitting is a phenomenon that occurs when a model that uses deep learning does not have enough information to accurately differentiate between the various classes that it needs to. This can cause the model to produce inaccurate results. Additionally, as the model level is increased, the point is reached at which the weight is saturated between the various classes that it needs to, a phenomenon known as overfitting occurs.

This accuracy point is achieved when the model has reached its optimal level. In addition, as the model level is increased, the point is achieved where the weight has reached its maximum potential. During the process of backpropagation, the gradient also disappears before it has a chance to reach the lower layers. The lowest layers do not obtain the most up-to-date information concerning the gradient.

It is not possible for a single configuration to address all these problems; however, the ensemble model that was proposed can. Because of the identity connections that exist between two convolutional layers, the residuals can be fed directly into almost all the other components of the model, which permits more rapid updates to the weights. The degree of detail that the model incorporates plays an important part in determining how well it can classify data and how well its findings can be applied to new situations.

The model that was proposed incorporates a total of four fundamental components, each of which is responsible for producing the optimal arrangement conceivable. Regardless of whether the level of model complexity is increased or decreased, the classification error will skyrocket after a certain benchmark has been reached. The model that was recommended has some advantages in terms of storage and safety; these advantages are since it does not store the iris patterns of particular people.

These benefits are a result of the device lack of an internal storage space for retinal patterns. It learns to differentiate between people based on model weights, which cannot be used to replicate eye patterns and are therefore not a threat to people right to privacy because they cannot be used to reproduce eye patterns.

The ensemble model that was proposed has a few advantages, but one significant drawback is that it is unable to account for variations in the total number of iris subjects that require classification. While this is a positive restriction, the ensemble model does have a few advantages.

The size of the output layer is predefined during model training and cannot be changed afterward, it is not feasible to dynamically add more iris subjects to discriminate. This is because the size of the output layer is predefined. This is since the dimensions of the output layer cannot be changed.

5. CONCLUSION

In this paper, the performance of the proposed ensemble method produces promising outcomes by reducing the amount of error to some degree. In addition to this, it is superior to the techniques that are currently considered to be state of the art. This is because the underlying ensemble model takes advantage of the rapid convergence provided by convolution and uses residual connections to solve the problem of vanishing gradients. The reason for this is that convolution offers quick convergence.

With the ultimate objective of improving feature discrimination, multiple methods of feature extraction are utilised to generate iris features from a variety of vantage points to accomplish this goal. The recommended technique is still able to recognise iris patterns, even though the datasets used in this research contain relatively large iris images that are not constrained in any way. This is because of the format of the procedure that was followed. It ensures that the method that is proposed is capable of handling noisy iris images in an appropriate fashion.

REFERENCES

[1] I.J. Jacob, "Capsule Network based Biometric Recognition System", *Journal of Artificial Intelligence*, Vol. 1, No. 2, pp. 83-94, 2019.

[2] R.M. Bolle, J.H. Connell, S. Pankanti, N.K. Ratha and A.W. Senior, "*Guide to Biometrics*", Springer, 2003.

[3] Iris Recognition, Available at: http://en.wikipedia.org/wiki/Iris_recognition, Accessed at 2022.

[4] Seifedine Kadry and Khaled Smaili, "A Design and Implementation of a Wireless Iris Recognition Attendance Management System", *Information Technology and Control*, Vol. 36, No. 3, pp. 323-329, 2007.

[5] M. Vatsa, R. Singh and A. Majumdar, "*Deep Learning in Biometrics*", CRC Press, 2018.

[6] S. Umer, A. Sardar and B.C. Dhara, "Person Identification using Fusion of Iris and Periocular Deep Features", *Neural Networks*, Vol. 122, pp. 407-419, 2020.

[7] H. Proenca and J.C. Neves, "Deep-Prwis: Periocular Recognition without the Iris and Sclera using Deep Learning Frameworks", *IEEE Transactions on Information Forensics and Security*, Vol. 13, No. 4, pp. 888-896, 2017

[8] N.V. Kousik and M. Saravanan, "A Review of Various Reversible Embedding Mechanisms", International *Journal of Intelligence and Sustainable Computing*, Vol. 1, No. 3, pp. 233-266, 2021

[9] S. Karthick and P.A. Rajakumari, “Ensemble Similarity Clustering Frame work for Categorical Dataset Clustering Using Swarm Intelligence”, *Proceedings of International Conference on Intelligent Computing and Applications*, pp. 549-557, 2021.

[10] Z. Zhao and A. Kumar, “A Deep Learning based Unified Framework to Detect, Segment and Recognize Irises using Spatially Corresponding Features”, *Pattern Recognition*, Vol. 93, pp. 546-557, 2019.

[11] B. Subramanian and S. Hariprasath, “Diabetic Retinopathy-Feature Extraction and Classification using Adaptive Super Pixel Algorithm”, *International Journal of Engineering and Advanced Technology*, Vol. 9, pp. 618-627, 2019.

[12] T.R. Sivapriya and V. Saravanan, “Automatic Brain MRI Mining using Support Vector Machine and Decision Tree”, *CiiT International Journal of Artificial Intelligent Systems and Machine Learning*, Vol. 3, No. 2, pp. 109-116, 2011.