Sclera Segmentation and Recognition for Spectacle Dataset

M.S Maheshan[1], C.K Roopa[1], Niharika M[1*], Nanditha M[1], Neha R[1], Kshamata Kaujalgi[1]

Department of Information Science & Engineering, JSS Science and Technology University, Mysuru, India[1]

maheshan@jssstuniv.in[1], ckr@sjce.ac.in[2], niharikamurulidhara@gmail.com[*3], nanditha908@gmail.com[4], Neha.ravi114@gmail.com[5], Kaujalgikshamata@gmail.com[6]

Abstract: Sclera, a connective tissue enveloping the eye, emerges as a novel biometric recognition method for human identification. The composition of blood vessels in the sclera proves ideal for biometric use—visible with ease, stable over time, and unique to each individual. This paper proposes sclera segmentation and recognition techniques tailored for individuals wearing spectacles. The SSV (Spectacle Sclera Vision) Dataset was meticulously created to address challenges introduced by eyewear, including reflections, distortions, and illumination variations. The study explores the unique characteristics of the sclera region, presenting a comparative analysis of traditional and neural network-based segmentation and recognition methods on the SSV Dataset. Notably, Linear SVC outperforms CNN in recognition, and UNET demonstrates superior sclera segmentation compared to OTSU. The findings provide a foundation for potential advancements in developing robust multi-class classification models for sclera biometrics in real-world scenarios. Future work involves further analysis, scalability testing, and exploration of diverse applications in ocular health and security systems.

Keywords: Sclera, Segmentation, OTSU, UNET, Linear SVC, CNN

I. INTRODUCTION

In recent years, the field of biometrics has undergone a significant surge in research and development, with a specific emphasis on elevating the precision and security of identification systems. Among the various avenues explored in this domain, there is a growing interest in sclera segmentation and recognition techniques, particularly tailored for individuals who wear spectacles. Given the substantial portion of the global population that relies on spectacles, optimizing biometric systems for this diverse demographic becomes imperative.

This paper presents a study on sclera segmentation and recognition methods, highlighting their application in the context of individuals wearing spectacles. The primary motivation is to tackle the distinctive challenges introduced by eyewear, including reflections, distortions, and variations in illumination [9], which often impede the accuracy of existing biometric solutions. The proposed research is the development and curation of the SSV (Spectacle Sclera Vision) Dataset—a diverse collection of scleral images featuring individuals with different eyewear types, lens prescriptions, and frames. This dataset helps in evaluating the effectiveness of sclera segmentation and recognition algorithms in real-world scenarios. Additionally, as biometrics relies on measurable physiological and behavioral traits [2] for identity verification, the focus extends to explores the unique characteristics of the sclera region. Traditionally dominated by iris and retina recognition methods, recent research is increasingly delving into additional ocular features that can complement iris-based identification [6], contributing to more secure and less-spoofable authentication schemes within the biometrics field.

The vein patterns in the sclera present a promising avenue for positive human identification, addressing the evolving needs of biometric authentication in a world where spectacle wearers form a significant user base.

Fig. 1. Structure of Eye and its Scleral Region

II. LITERATURE SURVEY

In recent years, there has been a discernible and notable shift in the scientific community's focus, diverting its attention from the traditionally explored domains of iris and retina recognition to the burgeoning field of sclera segmentation and recognition. This shift in emphasis reflects an evolving understanding of the potential benefits and applications of sclera-based biometric identification systems. Pathak and Tiwari [1] underscored the pivotal importance of several key factors that significantly influence the performance of sclera recognition systems. Among these factors are the development and utilization of efficient image segmentation algorithms, crucial for isolating and extracting relevant features from scleral images. Additionally, enhancing the visibility of blood vessels on the sclera surface emerged as a critical consideration, as it directly impacts the accuracy and reliability of recognition algorithms.

Expounding upon these foundational insights, Gokul Rajan and Vijayalakshmi [2] introduced a groundbreaking Modified Sclera Feature Extraction method. This novel approach leveraged the power of advanced machine learning techniques, particularly Convolutional Neural Networks (CNNs), to effectively address the inherent challenges posed by dynamic vessel patterns on the sclera. By incorporating sophisticated deep learning architectures, the proposed method achieved remarkable success, boasting an impressive 90.67% accuracy rate in identifying and classifying scleral features. This breakthrough not only highlights the potential of CNNs in tackling complex biometric recognition tasks but also underscores the importance of continuous innovation in algorithmic design for enhancing sclera recognition system performance.

Moreover, the exploration of sclera recognition has expanded beyond traditional single-view scenarios to encompass multi-view settings, where individuals' gaze directions vary significantly. Addressing the inherent challenges associated with such diverse viewing angles, a recent study [4] introduced a comprehensive approach that integrates advanced segmentation and representation techniques. This innovative methodology involves the implementation of a cascaded SegNet assembly for precise Region-Of-Interest (ROI) extraction, followed by the deployment of a dedicated CNN model, Sclera NET, for robust image representation. By effectively handling the complexities of multiview sclera recognition, this approach represents a significant step forward in broadening the applicability and robustness of scleral biometric systems across diverse real-world scenarios.

Furthermore, the quest for achieving higher levels of accuracy and reliability in sclera recognition has spurred research efforts towards exploring alternative recognition modalities. Notably, Sheng-Yu He and Fan [5] proposed a novel scheme centered around sclera vein recognition, leveraging a combination of traditional machine learning techniques, such as Support Vector Machine (SVM), and sophisticated feature extraction methods like Scale-Invariant Feature Transform (SIFT). With an impressive accuracy rate of 98.3%, this approach not only underscores the potential of vein-based biometrics but also hints at the possibility of further accuracy improvements through category expansion and refinement of feature extraction methodologies.

In parallel, Naqvi and Loh [6] introduced an innovative approach termed Sclera-Net, which focuses on semantic segmentation based on Information Maps (IM) and Non-Information Maps (NIM). By harnessing the power of semantic segmentation, this method demonstrated superior performance without the need for extensive preprocessing, thereby streamlining the recognition pipeline and enhancing efficiency. Additionally, Das and Harish [7] embarked on a crucial endeavour to establish a standard benchmark for eye recognition, with a specific focus on post-SSRBC 2016 developments in sclera segmentation. This comprehensive benchmarking initiative serves as a cornerstone for evaluating and comparing the efficacy of various sclera recognition methodologies, fostering collaboration and advancement within the research community.

A Modified Intuitionistic Fuzzy Clustering approach was proposed by M S Maheshan and B S Harish [10] for the effective segmentation of sclera images. The traditional fuzzy set assumes that the non-membership value is always the complement of the membership value. But in true sense, this assumption is not always correct because of hesitation. This method comprises two steps: Representation and Clustering. The proposed method alleviates the limitation of the existing IFCM based methods by employing a new intuitionistic complement generator. The proposed work also shows that for an effective recognition system of the sclera, segmentation plays a vital role.

III. METHODOLOGY

The process starts with capturing 248 images featuring 31 students, annotated for sclera using VGG annotation tool. After preprocessing, feature extraction is done using LBP method which lead to pattern recognition.

 https://doi.org/10.1051/itmconf/20246503014

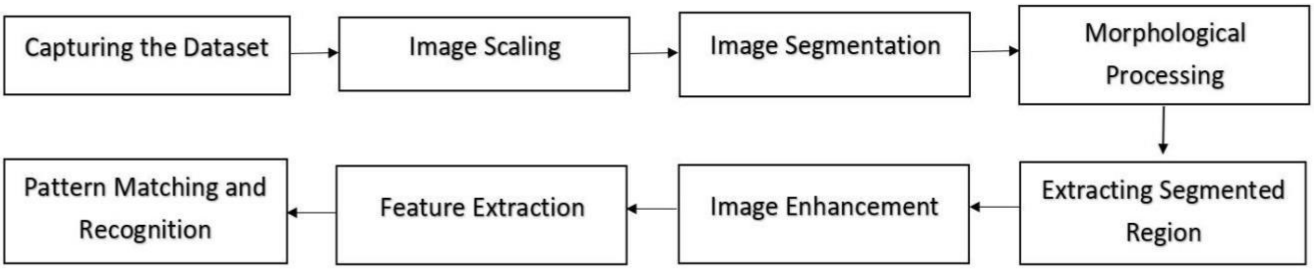

Fig. 2. Block Diagram of Methodology for Sclera Segmentation and Recognition

The dataset consists of 248 images of 31 students, captured using iPhone 14 Pro Max (48MP) and Oneplus 5T (36MP) cameras at 8 images per student. Ground truth masks, labeled "Sclera" and "non-sclera", were manually generated using the VGG annotating tool for all images.

Images were initially cropped and resized to 2250 x 1350 resolution, followed by down sampling for optimization purposes. Image segmentation techniques such as Otsu [1], Laplacian, Canny, and Roberts edge detection was employed to highlight features and accurately segment images based on pixel similarity.

Post-segmentation, morphological operations were applied to reduce noise and disturbances, leading to significant improvements in segmentation quality. The process involves converting images to grayscale, applying Otsu thresholding for binary mask creation, and performing bitwise AND operations to extract segmented regions.

Additionally, Histogram Equalization (HE) was applied to enhance image contrast and brightness. Local Binary Patterns (LBP) analysis and Random Forest/SVM classifiers were utilized for texture characterization and pattern recognition, respectively.

A. Dataset Creation – SSV (Spectacle Sclera Vision)

In the meticulous endeavour of capturing the dataset for this project, an extensive collection of 248 images featuring 30 people was systematically curated considering factors such as age, gender, lightning conditions, and spectacle type, ensuring representativeness, diversity, and quality while adhering to rigorous ethical and quality control standards. A total of 248 images, reflecting a thorough exploration of individual characteristics, were acquired through the advanced lens of an iPhone 14 Pro Max, boasting an impressive 48MP resolution. In order to make the dataset more challenging noise was introduced by capturing some images through a camera (OnePlus 5T) with less resolution that is 16MP. The seamless transfer of this visually rich SSV from the mobile device to a laptop. The organizational structure of the dataset revealed a deliberate arrangement, with each person contributing eight images neatly placed within individual folders, each folder distinguished by an alphabetically labelled class.

The annotation process of this SSV was a crucial phase, involving the utilization of the VGG annotating tool to generate ground truth masks manually. Creating ground truth masks manually involves meticulously outlining regions of interest within images to serve as reference data for training computer vision models. The specific class label assigned to denote the region of interest was "Sclera”. Notably, any remaining portions in the images were automatically classified as non-sclera. The generation of ground truth masks for all 248 images is done manually. To enhance the dataset's relevance and optimize it for the specific use case, a comprehensive preprocessing step was implemented. The process involved in preparing the images for analysis was meticulous and strategic. It began with the careful cropping of each image to extract the specific region of interest, namely the sclera, thereby ensuring that subsequent analyses would focus solely on the critical features relevant to sclera segmentation and recognition. Additionally, to standardize the dataset and facilitate consistent processing, all images underwent a resizing procedure, harmonizing them to a uniform resolution of 2250 x 1350 pixels. This attention to detail not only reflects a commitment to precision but also lays the groundwork for the subsequent phases of sclera segmentation and recognition. By standardizing the images in this manner, the effectiveness and applicability of the SSV (Sclera Segmentation and Recognition) system to the envisioned research objectives are optimized, setting a solid foundation for further analysis and experimentation.

Fig. 3. Raw Image from iphone

Fig. 4. Raw Image from oneplus 5

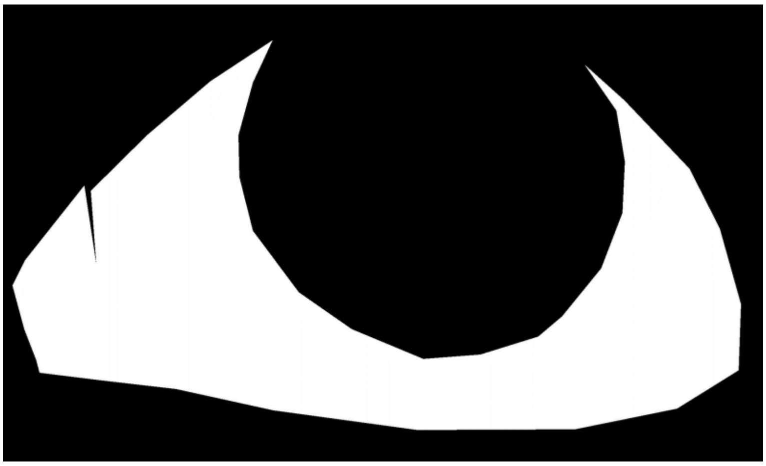

Fig. 5. Ground Truth Mask

B. Methodology

In image processing analysis, a comprehensive set of edge detection techniques, including Otsu, Laplacian, Canny, and Roberts, alongside a deep learning segmentation approach using UNET are applied. The results obtained from Otsu were used to train traditional models such as Linear SVC [9] and Random Forest. While the outcomes of UNET segmentation were employed to train a Convolutional Neural Network (CNN). These algorithms are suitable for datasets with complex feature spaces, such as those extracted from image textures or pixel intensities. Their robustness against overfitting, especially when dealing with noisy or high-dimensional data. Both provide some level of interpretability, Random Forest provides insights into feature importance, while Linear SVC reveals the decision boundary separating different classes.

For noise reduction in segmented images from Otsu, Opening is chosen over other techniques as it internally performs erosion followed by a dilation which removes small noise and preserves overall shape and structure of larger objects. On the other hand Closing only smoothens or connect irregularities. Subsequently, bitwise operations were performed between the original image and masks derived from both Otsu and UNET results to extract segmented regions:

The segmented regions were then converted to grayscale, and Histogram Equalization [1] was applied to enhance the sclera region.

For Otsu-segmented images, we employed the Local Binary Pattern (LBP) operator [8] to construct a feature vector representing the texture of the image.

The LBP formula is given by:

LBP (x_c, y_c) = Σ (i=0 to 7) s(i) * 2^i

Where,
(x_c, y_c): coordinates of the center pixels(i): comparison between the intensity of the center pixel and the neighbouring pixels.
i: index representing the neighbouring pixel position

Traditional models, specifically Linear Support Vector Classifier (SVC)[8] and Random Forest Classifier, were trained using the feature vectors obtained from LBP.

Linear SVC was configured with a random seed of 0 for reproducibility, a tolerance of 0.001, and a maximum of 10,000 iterations. The training process on pre-processed and scaled data demonstrated computational efficiency, completing in 0.18 seconds.

Random Forest Classifier used 50 decision trees, a fixed random seed of 42, and the same pre-processed and scaled training data, completing the training process in 0.172 seconds. The value 50 is chosen as greater number of decision trees reduces overfitting and increasing robustness to noise in the data. The selection of parameters such as the random seed, tolerance, and maximum iterations for Linear SVC, as well as the number of decision trees for Random Forest Classifier, is typically based on empirical experimentation and computational constraints. 42 is the default seed in many libraries, any integer value for random seed can reproduce the same results across different runs of the algorithm which helps in further debugging.

To train the Neural Network model, the mask obtained from UNET was overlaid on the eye image to extract the segmented region, and Histogram Equalization was applied for blood vessel enhancement.

The CNN model employed a Sequential architecture with three sets of convolutional layers (32, 64, and 128 filters), ReLU activation, and MaxPooling2D layers. Dropout layers (0.2 dropout rate) mitigated overfitting, and Dense layers applied ReLU and softmax activations. The CNN model employed a Sequential architecture with three sets of convolutional layers (32, 64, and 128 filters), ReLU activation, and MaxPooling2D layers. Dropout layers (0.2 dropout rate) mitigated overfitting, and Dense layers applied ReLU and softmax activations. ReLU introduces non-linearity while mitigating gradient issues. Its ability to accelerate convergence and perform well in image classification tasks makes it suitable for this CNN architecture with multiple convolutional and pooling layers. The model, compiled with categorical cross-entropy loss, Adam optimizer, and accuracy metric, underwent 50 epochs of training on a batch size of 128, evaluated against validation data.

1) Algorithm Overview:

Input: SSV Dataset

Output: Trained Model with Best Performance

The image processing workflow began with resizing the input images and applying various binarization and edge detection techniques. Otsu Thresholding was applied using cv2.threshold() from the OpenCV library, with the parameters set to the numpy array of the resized grayscale image, a threshold value of 125, a max value of 255, and the thresholding type set to cv2.THRESH_BINARY + cv2.THRESH_OTSU. This technique yielded the best results. Next, Canny Edge Detection was performed using cv2.Canny() with parameters including the resized grayscale image, a lower threshold of 80, and an upper threshold of 150. Additionally, Laplacian Edge Detection was applied using cv2.Laplacian() with the input as the resized grayscale image and the desired depth of the destination image set to -1. Roberts Edge Detection was implemented using filters.roberts() from the scikit-image library on the resized grayscale image. Given that Otsu Thresholding provided superior results, images processed with Otsu were further subjected to morphological processing techniques. The opening operation was applied using cv2.morphologyEx() from OpenCV with parameters including the resized grayscale image, the morphological operation type cv2.MORPH_OPEN, a 3x3 matrix of ones (np.ones((3, 3), np.uint8)), and one iteration. Similarly, the closing technique was applied using cv2.morphologyEx() with cv2.MORPH_CLOSE.

Subsequently, a neural network model, UNet, was employed on the downsampled images where the input images acted as features and the masks generated by the VGG Annotation tool served as labels. The UNet model's contracting path consisted of Conv2D layers with a kernel size of (3, 3) and ReLU activation functions, extracting features from the input images. Dropout layers with a rate of 0.1 were used to deactivate 10% of the neurons, mitigating overfitting. MaxPooling2D layers with a pool size of (2, 2) were utilized to downsample the feature maps. The outputs from each Conv2D layer were stored in variables (c1, c2, c3, c4, c5) for use in the expansive path. In the expansive path, the model used Conv2DTranspose layers for upsampling the feature maps by a factor of 2, followed by concatenation with the corresponding feature maps from the contracting path. For instance, u6 was derived by upsampling c5 using Conv2DTranspose with 128 filters, a (2, 2) kernel size, and a stride of (2, 2), maintaining the spatial dimensions through 'same' padding. This upsampled feature map was concatenated with c4, and the resulting tensor was processed through Conv2D layers with ReLU activation and a kernel size of (3, 3). Dropout layers with a rate of 0.2 were applied to further reduce overfitting.

The final UNet model was invoked with input parameters specifying the image height, width, and number of channels. To extract the segmented regions, the Otsu segmented mask was applied to the resized image using a bitwise operation with cv2.bitwise and (), combining the resized image and the segmented Otsu mask. The extracted segmented region was then enhanced using cv2.equalizeHist(), which equalized the histogram of the extracted region for better visual quality. Further LBP is used for Feature Extraction by providing HE enhanced images as input and the resulting feature vectors are utilized for training a recognition model.

IV. RESULTS

This study provides a comparative analysis of segmentation and recognition techniques on the SSV Dataset, using diverse performance metrics. Notably, the traditional recognition approach outperforms CNN for SSV.

Fig. 6. Otsu

Fig. 7. Laplacian

Fig 8. Roberts Edge

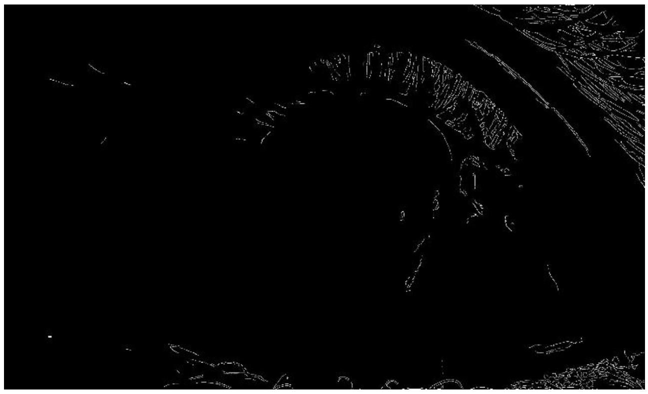

Fig. 9. Canny

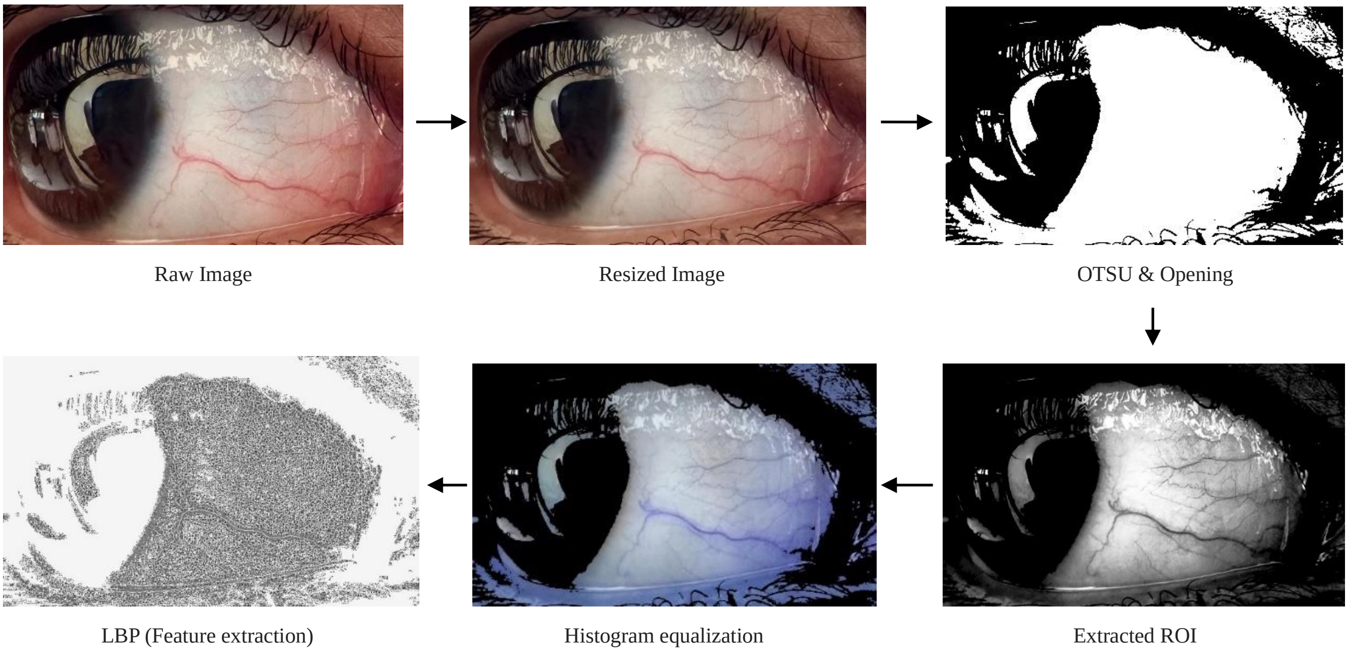

Fig. 10

TABLE 1- Traditional Segmentation algorithms result

Segmentation Technique	Jaccard Index (%)	Pixel Accuracy (%)	Time Complexity (sec)
Otsu	61.6	65.4	21.19
Laplacian	41.2	52.4	26.24
Canny	50.0	58.7	25.84
Roberts Edge	57.9	49.4	62.67

TABLE 2- Neural network Segmentation algorithm result

UNET	Ratios	Jaccard Index (%)	Pixel Accuracy (%)	Training Time (sec)	Prediction Time (sec)
	80:20	92.5	84.9	1477.68	2.12
	70:30	90.3	81.8	1313.55	3.97
	60:40	90.4	81.9	1414.16	8.69

The traditional OTSU segmentation achieved 65% accuracy, whereas UNET, a neural network approach, outperformed with 92% accuracy consistently across various ratios.

D. Discussion

In this section, we delve into the performance analysis of three distinct recognition techniques Linear SVM, Random Forest, and CNN. Our examination encompasses various training set ratios (60:40, 70:30, and 80:20) to comprehensively assess their efficacy in the realm of recognition tasks. Evaluating metrics such as accuracy, precision, recall, and training time sheds light on the comparative strengths and weaknesses of these methodologies. Linear SVM, Random Forest, and CNN, across various training set ratios (60:40, 70:30, and 80:20). Notably, Linear SVM achieves the highest accuracy at 80:20 ratio, reaching 55.6%, along with precision and recall of 64.7% and 55.1% respectively. Conversely, Random Forest's performance varies, with the highest accuracy of 31.0% at 80:20 ratio. CNN, however, exhibits consistently low accuracy across all ratios, with the highest at 18.05%. Moreover, CNN requires significantly longer training and prediction times compared to the other techniques, with training times ranging from 125.640 to 157.64 seconds. In summary, Linear SVM stands out with its superior accuracy and efficiency, while CNN lags behind due to its longer processing times and lower accuracy.

IV. CONCLUSION

In conclusion, the proposed work towards the implementation of sclera vein patterns of individuals wearing spectacles as biometric data. Notably, Linear SVC outperformed other recognition methods on the segmented images. Since the dataset is relatively small, linearly separable and contains patterns that can be effectively captured by a linear decision boundary. While, Random Forest and CNN requires larger datasets for training and more computational resources for optimization. They may also be prone to overfitting. Additionally, the deep learning segmentation technique UNET demonstrated enhanced results in sclera image segmentation compared to OTSU as the Images of individuals wearing spectacles involves intricate details and subtle variations in illumination, texture, and noise that UNET can better capture compared to OTSU, resulting in superior segmentation performance.

It is crucial to highlight that future analyses and experiments can be undertaken to delve deeper into the efficiency, scalability, robustness, and generalizability of the proposed system. This includes testing on larger datasets and evaluating performance under varying lighting conditions or occlusions. Overall, the results obtained underscore the effectiveness of Otsu segmentation [1] and Linear SVC for sclera segmentation and recognition in the spectacle dataset, serving as a foundation for potential advancements in developing models suitable for multiclass classification. This project lays the groundwork for the continued exploration and application of sclera biometrics, offering valuable insights for advancements in ocular health and security systems.

ACKNOWLEDGMENT

We express our gratitude to JSS Science and Technology University, Mysuru, for providing the necessary facilities for our project. Special thanks to our project guide, Dr. MS Maheshan, for his guidance and valuable suggestions throughout the project. We also appreciate the support and guidance from Dr. C K Roopa and Dr. B S Harish. Our thanks extend to the teaching and non-teaching staff of the Department of Information Science and Engineering for their support. Lastly, we are grateful to our family and friends for their constant support.

REFERENCES

[1] Pathak, R., Goswami, Y. C., & Tiwari, R. K. (2022). Sclera Segmentation and Sclera Vessel Enhancement. Journal of Positive School Psychology, 6(3), 7974-7980.

[2] Gokul Rajan V, Vijayalakshmi, S., & Sabharwal, M. (2021). Modified CNN based Feature Extraction for Sclera Recognition of Off Angular Images. Annals of the Romanian Society for Cell Biology, 25(3).

[3] Rajguru, M. S. T., Shedge, M. K. N., & Thosar, M. D. S. (2020). Robust Sclera Segmentation for Eye Gaze Detection and Sclera Recognition.

[4] Rot, P., Vitek, M., Grm, K., Emeršič, Ž., Peer, P., & Štruc, V. (2020). Deep sclera segmentation and recognition. In Handbook of vascular biometrics (pp. 395-432). Springer, Cham.

[5] He, S. Y., & Fan, C. P. (2019, March). SIFT features and SVM learning based sclera recognition method with efficient sclera segmentation for identity identification. In 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS) (pp. 297-298). IEEE.

[6] Naqvi, R. A., & Loh, W. K. (2019). Sclera-net: Accurate sclera segmentation in various sensor images based on residual encoder and decoder network. IEEE Access, 7, 98208-98227.

[7] Das, A., Pal, U., Ferrer, M. A., Blumenstein, M., Štepec, D., Rot, P.,... & Harish, B. S. (2017, October).SSERBC 2017: Sclera segmentation and eye recognition benchmarking competition. In 2017 IEEE international joint conference on biometrics (IJCB) (pp. 742-747). IEEE.

[8] Das, A., Pal, U., Ballester, M. A. F., & Blumenstein, M. (2014, December). A new efficient and adaptive sclera recognition system. In 2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM) (pp. 1-8). IEEE.

[9] Das, S., De Ghosh, I., & Chattopadhyay, A. (2021). An efficient deep sclera recognition framework with novel sclera segmentation, vessel extraction and gaze detecting. Signal Processing: Image Communication, 97, 116349.

[10] Maheshan, M. S., & Harish, B. S. (2021). A Modified Intuitionistic Fuzzy Clustering Approach for Sclera Segmentation. SN Computer Science, 2(4), 1-8.