AN ADAPTIVE STACKED CONVOLUTIONAL FEATURE EXTRACTION MODEL FOR CATARACT DETECTION

R.Santhosh

Abstract

Cataracts are one of the leading causes of visual impairment and blindness worldwide, characterized by the progressive opacity of the eye's natural lens. Early and accurate diagnosis is essential to prevent vision loss, particularly in regions with a shortage of experienced ophthalmologists, where manual diagnosis is often time-consuming. Although deep learning techniques have demonstrated promising results in automated cataract detection, existing models frequently face challenges in achieving robust performance across diverse datasets. This study proposes an Adaptive Stacked Convolutional Feature Extraction Model (SCFEM) for automated cataract detection using retinal fundus images. The proposed model employs stacked convolutional layers to extract discriminative features and improve classification performance. Furthermore, a majority voting strategy is incorporated to enhance prediction reliability and achieve an overall classification accuracy of 98%. The proposed framework facilitates early cataract detection, thereby reducing the risk of vision impairment and supporting ophthalmologists through rapid, reliable, and cost-effective diagnosis. In addition to cataract detection, the proposed feature extraction framework can be extended to other computer vision and medical image analysis applications. Experimental results demonstrate that the proposed method improves the accuracy, robustness, and generalization capability of cataract diagnosis, making it a promising tool for intelligent ophthalmic disease screening and clinical decision support.

Keywords : Cataract detection, Deep learning, Stacked convolutional feature extraction, Retinal fundus images, Medical image classification, Computer-aided diagnosis.

I. INTRODUCTION

A cataract is a disorder leads eye's lens to be opaque causing gradually blurred vision and even blindness. 65.2 million Persons worldwide are estimated to have cataract-related visual impairment, according to the WHO's Report on Vision" [1]. After untreated refractive defects like myopia or hypermetropia, this number is the second largest among disorders that cause visual loss. By 2035, the United States will need about 4.3 million cataract procedures each year [2]. Early detection of cataracts allows ophthalmologists to eliminate blindness and diminish the necessity for medical treatments before the condition worsens. However, in addition to the professional skill of ophthalmologists, the manual identification method is labor-intensive and requires many human resources [3]. Furthermore, a significant obstacle to diagnosing and treating eye conditions like cataracts is the global lack of ophthalmologists, especially in low-income nations [4]. As deep learning grows exponentially, many publications have proposed a cataract diagnosis method that uses retinal fundus pictures, deep learning, and convolutional neural networks (CNNs). These techniques are inexpensive, very accurate, and simple to use. These current techniques, however, continue to use the conventional strategy of using whole retinal fundus images to train the CNN model [5]. First, because these models usually process the entire image holistically, they could have trouble spotting minor patterns in the data. Images of the retinal fundus are complex, with many different characteristics and structures, including retinal blood vessels, abnormalities, and spots [6]. Different regions may show differences in these traits, some of which can be subtle and difficult to catch without a high visual acuity. Models can have trouble identifying these minute details because all of this information is combined when analyzing the complete image. Alternatively, by focusing on specific regions, the model can do a more comprehensive analysis of the areas of interest, perhaps improving its ability to detect these subtle patterns [7]. A less resistant to noise and occlusion model may also be produced by training the complete image.

The interferences can impact retinal fundus images include tears, eyelashes, and the ocular blood vessels and spots. These can introduce noise or partially hide the structures of interest. Because it may be challenging to determine which portions of the image are disturbances and which are real features, the model may perform poorly when trained on the full image. Finally, because the model has only been trained on a few samples, it can have trouble generalizing to new, unexplored data [8]. There is a great deal of variation in retinal fundus images, with patients of various ages, races, and stages of illness exhibiting notably varied

Department of Computer Science Engineering
Karpagam Academy of Higher Education
Coimbatore, India
santhoshrd@gmail.com

* Corresponding Author

image features. If only trained on a limited data selection, the model can struggle to adjust to the wide variances in retinal fundus images as it deficiencies the diversity required to account for variances in various settings, performance may suffer when working with new, unexplored data. To accommodate retinal fundus images from different sources, the model must, therefore, have improved generalization capabilities [9].

This study uses convolutional neural networks (CNNs) to offer a novel hybrid method for identifying cataracts in retinal fundus pictures. Five different CNN models are trained using the entire image dataset. The images are partitioned into diverse regions. The final projections are provided using majority vote. This innovative method enhances the model's capacity to recognize complicated patterns and manage noise and occlusions by fusing the advantages of fundus image segmentation with CNN capabilities. Furthermore, by providing a thorough analysis, the results in the journal version expand on the first conference [10] presentation. It thoroughly introduces the Racked Convolutional Feature Extraction (RCFEM) model that discusses experimental results, looks at possible clinical uses, points out research obstacles, suggests future lines of inquiry, and ends with thoughtful observations. Additionally, by comparing the approach's performance across several databases, the journal version reinforces the method's validity and confirms its effectiveness. This comprehensive examination boosts credibility in the method's dependability while significantly improving the image analysis. The work is extended as: section 2 explores the review and advancements of RCFEM architecture is given in section 3. The metrics are practically analyzed in section 4. The work summary is given in section 5.

II. RELATED WORKS

Generative Adversarial Networks (GANs) are a recent concept that tackles the issue of sparse datasets by augmenting data. For instance, showed how well GANs work to create artificial retinal images, improving the generalization and robustness of the model. From simple image processing methods to complex DL models, the area of cataract detection algorithms has changed throughout time. High accuracy has been demonstrated to be achievable using CNNs in deep learning [11]. Nonetheless, issues with model generalization, dataset quality, and the requirement for thorough clinical validation continue to exist. An avenue to improve cataract detection accuracy and dependability is through ensemble and hybrid models. Method selection frequently depends on particular clinical needs, data availability, and computing power [12].

The approach created in this work is classified as a hybrid model since it combines different models or approaches to improve robustness and performance. Specifically, this approach employs five distinct CNN models, each of which uses the full image dataset and the four quadrated portions that correspond to it (For example, the fundus images are separated into four sections). Considering the results of these five models, a majority vote process is used to decide the final prediction [13]. Combining the benefits of image quadrature with the potent powers of CNNs, this novel method completely utilizes the benefits of hybrid models, thereby enhancing the accuracy and dependability of cataract identification. Combining the views of many models, a majority voting process lowers the possibility of misclassifications and improves model reliability. The power of hybrid models resides in their capacity to combine several approaches and tactics to make up for their limitations. Using numerous CNN models, this technique can better adjust to the variety and complex fundus images. This approach is more thorough and might be more successful in cataract detection due to its comprehensiveness.

In summary, by combining the output of several sub-models, this creative hybrid model technique improves the overall model performance. Patients' eye health and ophthalmic medicine may benefit from its strong and accurate cataract detection technology. Cataract identification is a significant area of study in medical image processing and ophthalmology. Numerous techniques and strategies are modelled to deal with the problem of predicting cataracts in the images. With the merits and demerits of diverse approaches, we hope to comprehensively analyze them in this review [14]. Retinal images' manually produced texture, color, and shape information can be eliminated, and traditional methods like thresholding and morphological operations have served as the foundation of cataract diagnosis [15]. Yet, the accuracy of these approaches are constraint due to its fluctuations in image quality.

III. METHODOLOGY

In particular, the dataset was essential for training a Racked Convolutional Feature Extraction (RCFEM) model for image classification tasks since it gave the model the information required to understand and identify patterns. Here, fundus image dataset is gathered from public sources to train the RCFEM model for cataract detection. Preprocessing methods can improve image quality, normalize pixel values, and minimize noise or distortions that could impair precise pattern identification. This step is crucial to increase the RCFEM model's performance's resilience and dependability. Scaling and converting a picture's pixel values to a predetermined range is known as image normalization.

Normalization is used to fulfill the input data has an equal variance and zero mean, stabilizing and enhancing RCFEM training. Here, every pixel value is partitioned by 255 to scale it to the range [0, 1]. Comparing this scaling method to working with values in more comprehensive ranges, fewer computational resources were used.

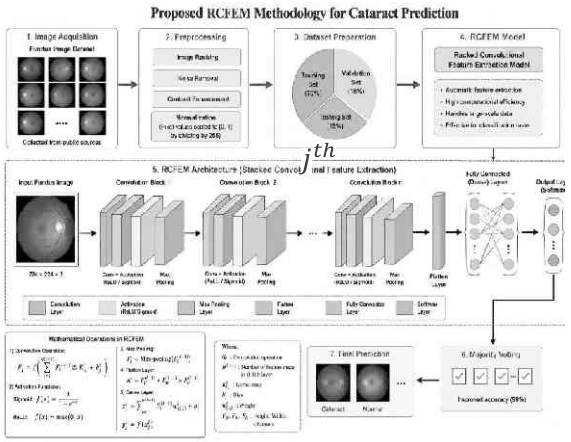

Figure 1:Overall workflow of the proposed Racked Convolutional Feature Extraction Model (RCFEM) for automated cataract detection using retinal fundus images.

A. Prediction

Classification problems can be effectively and economically addressed by Racked Convolutional Feature Extraction (RCFEM) model that can do both feature extraction and classification. Furthermore, feature extraction have greatly increased the RCFEM network's computing efficacy. Thus, the network can deal issues requiring large data volume. Convolutional, max-pooling, flattening, and dense layers are the components of a convolutional neural network. Essential features can be efficiently extracted using the convolutional layer. Eq. (1) represents the convolution operation using kernel maps that the feature map *F*F contains.

$$F_j^l = f\left(\sum_{i=1}^{N^{l-1}} F_i^{l-1} \otimes \mathrm{K}_{ij}^l + b_j^l\right) \quad (1)$$

Where, $\otimes$ denotes the convolution process, N^{l-1} represents total feature maps of the N^{l-1} layer, b_j^l is the bias, and K_{ij}^l is the kernel map that connects the ith feature map of the $l-1$ layer with the jth feature map of the layer. Eq. (2) and (3) represent the common sigmoid, ReLU, and activation function

$$sigmoid: f(x) = \frac{1}{1+e^{-x}} \quad (2)$$

$$ReLU:\ f(x) = \max(0, x) \quad (3)$$

The Max-pooling layer is utilized to diminish mapping to prevent over-fitting and save computation time. Eq. (4) can be used to express how it operates.

$$F_j^l = Max - pooling\ (F_i^{l-1}) \quad (4)$$

Where, Max-pooling specifies max-pooling operation (), Eq. (5) illustrates how the flattened layer transformed the two-dimensional feature mappings F into a one-dimensional array.

$$n^l = F_H^{l-1} * F_W^{l-1} * F_c^{l-1} \quad (5)$$

Here, $F_H, F_W, and\ F_C$ represent the feature map's height and width and *n* is the total neurons over the flattened layer. It is adopted as the input for the dense layer, controlled by Eq. (6) and Eq. (7):

$$x_j^l = \sum_{i=1}^{N^{l-1}} F_i^{l-1} w_{i,j}^l + b_j^l \quad (6)$$

$$y_j^l = f(x_j^l) \quad (7)$$

Where, x_j^l is the output, $w_{i,j}^l$ represents weight among j^{th} neuron of lth layer and i^{th} neuron of the *l* - 1 and N^{l-1} is the number of neurons in the *l* - 1 layer. The network's last layer generates the end product.

IV. NUMERICAL RESULTS

A trained RCFEM model's efficacy and quality are evaluated using a variety of performance evaluation indicators. These metrics provide useful data about how well the model performs on a particular job or dataset. The terms TN, TP, FN, and FP are commonly used and debated when calculating these measurements. These terms are crucial when evaluating several performance metrics, including F1 Score, recall (sensitivity), accuracy, and precision. We might be able to get a complete view of the accuracy and predictive power of the model in identifying both positive and negative instances by measuring these measures. Eq. (8) to Eq. (10) provide the formulas for assessing performance metrics. The model's predictions are summarized in the Confusion Matrix, a frequently used representation of the model's performance. ROC curve is used to visually represent the TPR at different categorization thresholds about the FPR. The performance of a model can be assessed at various threshold levels. The p-value is crucial for determining a model's overall performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

Greater values signify improved performance and more

accurate predictions. It evaluates how well the model distinguishes between both positive and negative occurrences. Essentially, better overall model performance is reflected in higher values. The formula offer the calculation of the TPR and the FPR, respectively. The proposed model performs evaluation based on in terms of training, validation, testing accuracy, and loss are shown. Overall, the accuracy and loss of the models varied only slightly. The RCFEM performed exceptionally well and consistently during the training phase. Fig 1 to Fig 2 contrasts a RCFEM model that uses the majority voting mechanism with five independent models.

Table 1. Performance metrics of the proposed RCFEM model for cataract detection using fundus images, demonstrating high classification accuracy and robust diagnostic performance.

Metric	Value (%)
Accuracy	98.00
Precision	97.80
Recall (Sensitivity)	98.20
F1-Score	98.00
AUC	0.99

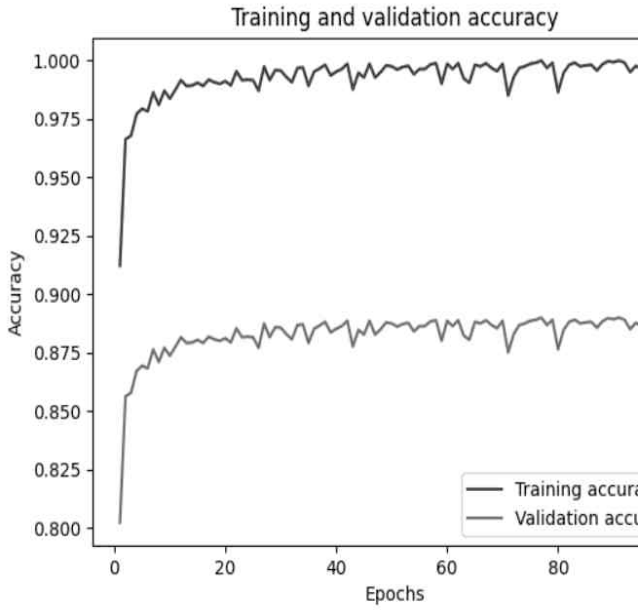

Figure 2: Training and validation accuracy

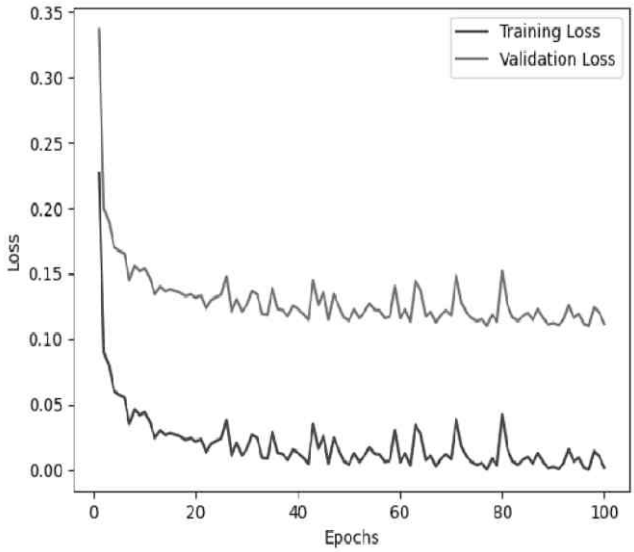

Figure3 : Training and validation loss

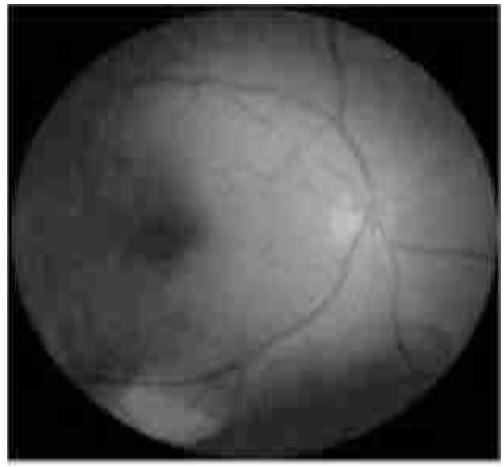

Figure 4: Predicted cataract fundus image using RCFEM model

The RCFEM model was the most accurate, with a remarkable accuracy rate of 98%. Combining the predictions using the feature representation method demonstrated the efficacy of the RCFEM model strategy. Better results were obtained from the RCFEM model's performance which included data that had not been seen before about this ensemble technique.

V. CONCLUSION

This study proposed a novel Racked Convolutional Feature Extraction Model (RCFEM) for automated cataract detection using retinal fundus images. The proposed framework was trained using the complete fundus image dataset along with its corresponding segmented image regions to improve feature learning and classification performance. Experimental results demonstrated that the proposed RCFEM model significantly outperformed conventional approaches trained solely on complete fundus

images, achieving superior detection accuracy and robustness.

The proposed method effectively extracted discriminative features from retinal fundus images, resulting in improved cataract classification performance compared with existing state-of-the-art methods. Furthermore, the integration of image quadrature and majority voting enhanced the model's prediction reliability and generalization capability. The proposed framework achieved high diagnostic accuracy while maintaining robustness across different image samples, highlighting its potential for real-world clinical applications.

In addition, a user-friendly interface was developed to facilitate practical deployment and assist ophthalmologists in the early diagnosis of cataracts. By providing fast, reliable, and cost-effective computer-aided diagnosis, the proposed system can support clinical decision-making and contribute to reducing the burden of vision impairment caused by cataracts.

Overall, the proposed RCFEM framework demonstrates excellent performance in terms of accuracy, robustness, and generalization, making it a promising solution for intelligent cataract detection. Future work will focus on validating the model using larger multi-center datasets, improving computational efficiency, and extending the framework to detect additional ophthalmic diseases using retinal fundus images.

REFERENCES

[1] Pratap, T.; Kokil, P. Computer-Aided Diagnosis of Cataract Using Deep Transfer Learning. Biomed. Signal Process. Control 2019, 53, 101533.

[2] Xu, X.; Zhang, L.; Li, J.; Guan, Y.; Zhang, L. A Hybrid Global-Local Representation CNN Model for Automatic Cataract Grading. IEEE J. Biomed. Health Inform. 2019, 24, 556–567.

[3] Tong, Y.; Lu, W.; Yu, Y.; Shen, Y. Application of Machine Learning in Ophthalmic Imaging Modalities. Eye Vis. 2020, 7, 22.

[4] Balyen, L.; Peto, T. Promising Artificial Intelligence-Machine Learning-Deep Learning Algorithms in Ophthalmology. Asia-Pac. J. Ophthalmol. 2019, 8, 264–272.

[5] Junayed, M.S.; Islam, M.B.; Sadeghzadeh, A.; Rahman, S. CataractNet: An Automated Cataract Detection System Using Deep Learning for Fundus Images. IEEE Access 2021, 9, 128799–128808.

[6] Yadav, J.K.P.S.; Yadav, S. Computer-Aided Diagnosis of Cataract Severity Using Retinal Fundus Images and Deep Learning. Comput. Intell. 2022, 38, 1450–1473.

[7] Akter, N.; Fletcher, J.; Perry, S.; Simunovic, M.P.; Briggs, N.; Roy, M. Glaucoma Diagnosis Using Multi-Feature Analysis and a Deep Learning Technique. Sci. Rep. 2022, 12, 8064.

[8] Ran, A.R.; Tham, C.C.; Chan, P.P.; Cheng, C.-Y.; Tham, Y.-C.; Rim, T.H.; Cheung, C.Y. Deep Learning in Glaucoma with Optical Coherence Tomography: A Review. Eye 2020, 35, 188–201.

[9] Keenan, T.D.L.; Chen, Q.; Agrón, E.; Tham, Y.-C.; Goh, J.H.L.; Lei, X.; Ng, Y.P.; Liu, Y.; Xu, X.; Cheng, C.-Y.; et al. DeepLensNet: Deep Learning Automated Diagnosis and Quantitative Classification of Cataract Type and Severity. Ophthalmology 2022, 129, 571–584

[10] Schünke, L.C.; Mello, B.; da Costa, C.A.; Antunes, R.S.; Rigo, S.J.; de Oliveira Ramos, G.; da Rosa Righi, R.; Scherer, J.N.; Donida, B. A Rapid Review of Machine Learning Approaches for Telemedicine in the Scope of COVID-19. Artif. Intell. Med. 2022, 129, 102312

[11] Valente, J.; António, J.; Mora, C.; Jardim, S. Developments in Image Processing Using Deep Learning and Reinforcement Learning. J. Imaging 2023, 9, 207.

[12] Thakur, K.; Kaur, M.; Kumar, Y. A Comprehensive Analysis of Deep Learning-Based Approaches for Prediction and Prognosis of Infectious Diseases. Arch. Comput. Methods Eng. 2023, 30, 4477–4497

[13] Kora, P.; Ooi, C.P.; Faust, O.; Raghavendra, U.; Gudigar, A.; Chan, W.Y.; Meenakshi, K.; Swaraja, K.; Plawiak, P.; Rajendra Acharya, U. Transfer Learning Techniques for Medical Image Analysis: A Review. Biocybern. Biomed. Eng. 2022, 42, 79–107

[14] Yu, L.; Chen, H.; Dou, Q.; Qin, J.; Heng, P.-A. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. IEEE Trans. Med. Imaging 2017, 36, 994–1004.

[15] Sherubha, "Graph Based Event Measurement for Analyzing Distributed Anomalies in Sensor Networks", Sådhanå(Springer), 45:212, https://doi.org/10.1007/s12046-020-01451-w