

LBPSCN: Local Binary Pattern Scaled Capsule Network for the Recognition of Ocular Diseases

Mavis Serwaa¹, Patrick Kwabena Mensah², Adebayo Felix Adekoya³, Mighty Abra Ayidzoe⁴

Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana^{1,2,4}

Department of Computer Science, Sunyani Technical University, Sunyani, Ghana¹

Department of Computing and Information Sciences, Catholic University of Ghana, Sunyani, Ghana³

Abstract—Glaucoma and cataracts are leading causes of blindness worldwide, resulting in significant vision loss and quality of life impairment. Early detection and diagnosis are crucial for effective treatment and prevention of further damage. However, diagnosis is challenging, especially when intraocular pressure is low or cataracts are present. Deep learning algorithms, particularly Convolutional Neural Networks (CNNs), have shown promise in detecting eye diseases but require large training datasets to achieve high performance.. To address this limitation, this work proposes a modified Capsule Network algorithm with a novel scaled processing algorithm and local binary pattern layer, enabling robust and accurate diagnosis of glaucoma and cataracts. The proposed model demonstrates performance comparable to state-of-the-art methods, achieving high accuracy on combined, cataract-only, and glaucoma-only datasets (94.32%, 96.87%, and 95.23%, respectively). This work introduces enhanced feature extraction and robustness to illumination variations, addressing critical limitations of existing methods.. The proposed model offers a promising tool for ophthalmologists and glaucoma specialists to accurately diagnose glaucoma and cataract-compromised eyes, potentially improving patient outcomes.

Keywords—Glaucoma; cataracts; capsule network; convolutional neural network

I. INTRODUCTION

Glaucoma is a leading cause of blindness worldwide, resulting from progressive optic nerve degeneration. Unlike cataracts, which can be reversed through surgery, glaucoma damage is irreversible. Early detection can halt further damage, but diagnosis is challenging, especially when intraocular pressure is low. By the year 2040 with approximately 111.8 million people susceptible to ocular diseases[1], developing intelligent algorithms for telemedicine-based screening and diagnosis is crucial for early detection and prevention of vision loss.

Deep learning algorithms, particularly Convolutional Neural Networks (CNNs), have shown promise in detecting eye diseases. CNNs require large training datasets to achieve high performance and prevent over-fitting on data. However, medical datasets are limited, smaller in size, and imbalanced. Data augmentation techniques, which are time-consuming and may miss critical image poses, are adopted as a fallback approach to increase data size and circumvent data-induced overfitting issues. Several approaches have been proposed to address the limitations of CNNs. Capsule Networks are a prominent algorithm that captures the characteristics of CNNs and addresses their data-induced challenges. The introduction of this innovative concept ignited a wave of interest among researchers from diverse fields, inspiring them to investigate

its capabilities and push its boundaries. Capsule Networks (CapsNets) are equivariant and adaptable to smaller datasets. However, their encoder network is weak [2], and feature processing is insufficient.

This work proposes a modification to the encoder network drawbacks of the Capsule network algorithm and adapt it to diagnose glaucoma, cataract, non-glaucoma, and non-cataract images. A feature enhancement algorithm termed scaled processing algorithm is proposed. The technique applies weights to the feature maps. Softmax activation function is applied to the scaled feature maps to enhance contrast. The model is made computationally efficient and robust to illumination variations by the incorporation of a local binary pattern layer (LBP). The proposed model performs comparably well with state-of-the-art methods and can assist ophthalmologists and glaucoma specialists in effectively diagnosing cataracts and glaucoma-compromised eyes. The contributions of the paper are as follows:

- **Enhanced feature extraction:** This work introduces a novel scaled processing algorithm, which significantly enhances feature maps, leading to improved recognition accuracy and addressing a critical limitation of existing methods
- **Robustness to illumination variations:** The proposed model incorporates a local binary pattern layer (LBP), ensuring robustness to illumination variations, a common challenge in fundus image analysis, and thereby improving the reliability of diagnosis.
- **Accurate diagnosis of glaucoma and cataract:** The proposed Capsule Network model demonstrates comparable performance to state-of-the-art methods, offering a promising tool for ophthalmologists and glaucoma specialists to accurately diagnose glaucoma and cataract-compromised eyes, potentially improving patient outcomes.

The rest of the paper is organized as follows: Section II presents related works in glaucoma and cataract detection. Section III describes the proposed methods, model, dataset, and experimental settings. Section IV presents the results and discussion, and Section V concludes the paper.

II. RELATED WORKS

The detection and diagnosis of glaucoma and cataracts have been extensively researched in the field of medical image analysis. Various Deep Learning approaches have been proposed to

improve the accuracy and efficiency of diagnosis, leveraging advancements in convolutional neural networks (CNNs) and capsule networks (CapsNets). In recent years, several studies have explored the application of CNNs and CapsNets to fundus images, optical coherence tomography (OCT) scans, and other retinal imaging modalities. In the domain of CNNs, Oguz et al. [3] proposed a CNN-based hybrid model for the recognition of glaucoma disease. The hybrid trait was achieved by the infusion of Adaboost into the CNN model. The proposed model combines and processes deep features and machine learning features extracted from fundus images. The proposed model achieved 92.96% accuracy, 93.75% F1-score, and an AUC of 0.928 when experimented on the ACRIMA dataset. Velpula and Sharma [4] adopted the approach of exploring pre-trained CNN models (thus, ResNet50, AlexNet, VG19, DenseNet-201, and Inception-ResNet-v2) on glaucoma datasets and developed a fusion mechanism to combine and weight the results of the pre-trained models. These CNN models were explored on four datasets (thus, RIM-ONE, ACRIMA, Harvard Dataverse, and Drishti) and achieved 99.57%, 85.43%, 90.55%, and 94.95% recognition accuracies on the ACRIMA, Harvard Dataverse, RIM-ONE, and Drishti datasets, respectively. Shoukat et al. [5] adopted the ResNet-50 architecture and fine-tuned it to detect glaucoma. Data augmentation techniques were adopted to increase and develop diverse orientations of the fundus images. The proposed model analyses patterns in the retinal images that are not considered for diagnosis by medics. The proposed model achieved 98.48% detection accuracy. In the domain of CapsNets, [1] applied the original capsule network with dynamic routing on a dataset containing retina images of glaucoma. The CapsNet model attained 90.90%, 86.64%, 90.59%, and 0.904 on the accuracy, recall, precision, f1-score, and kappa index, respectively. Gaddipati et al. [6] modified the capsule network with dynamic routing to make it suitable for processing 3D optical coherence tomographic images. The proposed model comprised 3D convolution layers, batch layers, and leaky rectified activation functions. Two self-collected eye datasets were combined and randomly split into three categories for the experimental training. The images in the combined dataset were resized to 64x64x128. The proposed model attained 0.89, 0.96, 0.94, and 0.973 values on sensitivity, specificity, accuracy, and area under curve. Ayidzoe et al.[7] proposed an enhanced capsule network. The capsule network's encoder layer was optimized by the proposal and development of a feature enhancement algorithm termed feature amplification. The proposed method enables the proposed model to focus on relevant features. The proposed model was trained on the ODIR and an eye disease dataset. The glaucoma and cataract classes achieved (0.796, 1.00) and (0.818 and 0.987) values on precision and specificity, respectively. Capsule Network have not been fully explored for eye disease recognition as it is a trending area and relatively new compared to CNNs.

III. MATERIAL AND METHOD

This section presents the capsule Network algorithm, the local binary pattern algorithm, the proposed model, the dataset description, and the experimental setup.

A. Capsule Network

The concept of capsules was proposed by Hinton et al. [8] and its fully-fledged concept was modified and implemented by Sabour et al. [9]. These two works presented the idea of Capsule Network in an informative and promising way of handling image classification. This novel concept sparked widespread interest among researchers across various domains, prompting them to explore its potential.

A capsule refers to a vector comprising the Properties of an object's part. A collection of capsules forms a capsule layer. Stacks of Convolutional layers, Capsule layers (thus, the primary and class capsule layers), and fully connected layers make up a Capsule Network. The Convolutional layers and Capsule layers constitute the encoder network whereas the fully connected layers constitute the decoder layer. The length of the vector of a capsule signifies the presence or absence of the features of an entity. The values in the vectors are generated by the neurons in the convolutional layers. In the primary capsules, the vectors of the capsules t_i , are transformed via a transformation matrix, w_{ij} . This procedure encodes some characteristics (such as rotation, scaling, and many more) of capsules in the class capsule layer to the capsules in the primary capsule layer (PC). The capsules in the primary capsule layer can be seen as children capsules whereas the capsules in the class capsule layer can be seen as parent capsules. The transformation procedure produces vectors termed prediction vectors and these prediction vectors are computed as follows (Eq. 1):

$$\hat{x}_{ij} = t_i * w_{ij} \quad (1)$$

Each prediction vector (thus, the modified capsules in the PC layer) actively searches for its parent capsules during each iteration. To establish a linkage, coupling coefficients are calculated. The linkage shows that a child capsule's properties can be found in the feature space of a parent capsule. This procedure is termed coupling and its coefficients are computed as follows:

$$p_{ij} = \frac{\exp(y_{ij})}{\sum_k \exp(y_{ik})} \quad (2)$$

Where y_{ij} are learned log prior probabilities which are updated during training. The coupling coefficients p_{ij} , are applied to the prediction vectors \hat{x}_{ij} , to establish a relationship between the child and the parent capsules. This is computed as follows:

$$r_j = \sum_{a=1}^A p_j \hat{x}_{aj} \quad (3)$$

The length of the modified prediction vectors signifies two capsules share similar properties or not. If the length of the modified prediction vectors, is close to one, it means a child capsules in the primary capsule layer share similarity with a parent capsule and in the opposite condition, the length of the prediction vectors is squashed to zero. This squashing procedure is computed as follows:

$$m_j = \frac{\|r_j\|^2}{1 + \|r_j\|^2} \frac{r_j}{\|r_j\|} \quad (4)$$

The softmax activation function is applied to the final outputs to convert the prediction results into probabilities. This is the summary of what happens in the encoder network. The decoder network on the other hand is designed to produce reconstructed images of the input.

B. Local Binary Pattern

The local binary pattern (LPB) algorithm [10] works by extracting key point textural information from images. These key point features are extracted via the thresholding of neighboring pixels of each pixel in the subject image. The equivalent binary versions of the neighboring pixels are computed. The concept of LBP was adopted in this work to achieve the proposal of a lightweight Capsule network algorithm as LBP is not a computationally intensive algorithm and has a novel way of collecting chrominance data from images. The computation of an LBP descriptor follows four steps:

- For every pixel (x, y) in an image, choose F neighboring pixels at radius G.
- Calculate the strength difference of the current pixel (x, y) with the F neighboring pixels.
- Threshold the strength difference, such that all the negative differences are apportioned as 0 and all the positive differences are apportioned as 1, forming a bit vector.
- Convert the F-bit vector to its corresponding decimal value and replace the strength value at (x, y) with this decimal value.

Fig. 1 shows an example of a computed LBP version of an image. The LBP descriptor is given as this (Eq. 5):

$$LBP(F, G) = \sum_{i=0}^{N-1} 2_i s(i_i - i_c) \quad (5)$$

where F is given as the neighboring pixels, G is given as the radius, i_i and i_c denotes the intensity of the current and neighboring pixels respectively. s is a sign function defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

C. Scaled Processing Algorithm

To boost the performance of the proposed model, a feature enhancement algorithm was adopted and modified. We modified the feature amplification algorithm from the work of [7]. For the amplification procedure, the weight of 2 was used to multiply the feature maps. To make the model focus on the relevant areas, we passed the amplified features through a softmax activation function. The proposed method makes bright pixels brighter and dark pixels darker consistent with the observation of Nguyen et al. [11]. The proposed method boosts

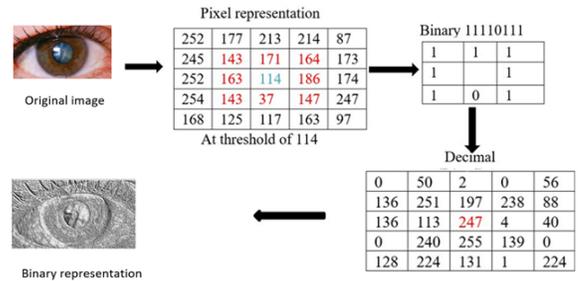


Fig. 1. Creation of binary representation of an image using LBP.

the performance of the proposed model compared to the state-of-the-art. The feature enhancement algorithm termed scaled amplification is shown in algorithm 1. Fig. 2 shows the pixel intensity of both the raw input image and its scaled version. As shown in the pixel intensity plot of the scaled image, the scaled processing algorithm appears to shift the model's focus to relevant features, evidenced by reduced pixel intensity around smaller pixel values. The mathematical formula of the proposed scaled processing algorithm is presented as:

$$p_{a,b}^k = \frac{s * (exp(l_{a,b}^k))}{s * \sum_{k=1}^{K-1} (l_{a,b}^k)} \quad (6)$$

Algorithm 1 Scaled Processing Algorithm

1. Input: $L_{a,b}^K = l_{a,b}^0, l_{a,b}^1, \dots, l_{a,b}^{K-i} \triangleleft$ feature maps
2. Output: $p_{a,b}^k$
- To preprocess and improve the contrast, $\forall L_{a,b}^K$
3. for feature map k in $L_{a,b}^K$ do
4. $t_{a,b}^k = l_{a,b}^k * s$ where $1 > s \leq 5$
5. $p_{a,b}^k = \text{softmax}(t_{a,b}^k)$
6. end
7. return $p_{a,b}^k$

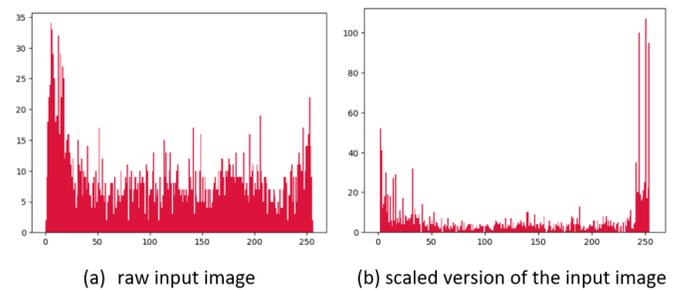


Fig. 2. The pixel intensity of the non-scaled input image and its scaled version (scaled processing algorithm output).

D. Proof of the Scaled Processing Algorithm

Lemma 1: The scaling operation in step 3, $t_{a,b}^k = l_{a,b}^k * s$, where $1 < s \leq 5$, amplifies the feature values while preserving their relative proportions.

Proof: Let $t_{a,b}^k$ be the original feature value and s be the scaling factor. Then, $t_{a,b}^k * s$ is the scaled feature value. Since $1 < s \leq 5$, we have:

$$t_{a,b}^k * s > t_{a,b}^k$$

This implies that the scaling operation amplifies the feature values. Moreover, since s is a constant, the relative proportions between the feature values are preserved.

Lemma 2: The softmax function in step 4, $p_{a,b}^k = \text{softmax}(t_{a,b}^k)$, normalizes the scaled feature values to a probability distribution.

Proof: The softmax function is defined as:

$$\text{softmax}(x) = \frac{\exp(x)}{\sum \exp(x)}$$

Since $t_{a,b}^k$ is the input to the softmax function, we have:

$$p_{a,b}^k = \frac{\exp(t_{a,b}^k)}{\sum \exp(t_{a,b}^k)}$$

This implies that the softmax function normalizes the scaled feature values to a probability distribution, where each value represents the probability of the corresponding feature being important.

Theorem: The proposed algorithm enhances the feature maps and enables the Capsule Network model to attain high accuracy.

Proof: By Lemma 1, the scaling operation amplifies the feature values while preserving their relative proportions. By Lemma 2, the softmax function normalizes the scaled feature values to a probability distribution. Therefore, the proposed algorithm enhances the feature maps by amplifying important features and suppressing unimportant ones. This leads to improved accuracy in the Capsule Network model.

The weight "s" must be between 1 and 5, else there will be noise in the feature maps

Lemma 3: If $s \geq 5$, then the scaling operation $t_{a,b}^k = t_{a,b}^k * s$ will produce noise in the feature maps.

Proof: Let $t_{a,b}^k$ be the original feature value and $s \geq 5$ be the scaling factor. Then, we have:

$$t_{a,b}^k * s \geq 5 * t_{a,b}^k$$

Since $t_{a,b}^k$ is a feature value, it is typically normalized to have a small magnitude (e.g., between 0 and 1). When we multiply it by $s \geq 5$, the result is a very large value, which can cause numerical instability and produce noise in the feature maps. Furthermore, when we apply the softmax function to these scaled values, the noise will be amplified, leading to a

distorted probability distribution. This distortion can cause the Capsule Network model to produce inaccurate results.

Theorem: If the weight s is set to 5 or more, there will be noise produced in the feature maps, leading to inaccurate results in the Capsule Network model.

Proof: By Lemma 3, if $s \geq 5$, the scaling operation will produce noise in the feature maps. This noise will be amplified by the softmax function, leading to a distorted probability distribution. Therefore, the proposed algorithm will produce inaccurate results if the weight s is set to 5 or more.

E. Proposed Model

Fig. 3 presents the proposed model. It comprises two scaled layers, one local binary pattern layer, three convolutional layers, a primary capsule layer, a class capsule layer, and three fully connected layers. The local binary layer uses a filter of size 3x3. The first and second convolutional layer has 64, 3x3 filters. The second convolutional layer has 256, 3x3 filters. The capsules in the primary capsule have a dimension of eight whereas the capsules in the class capsule have a dimension of 2. The first, second, and third fully connected layers have 512, 1024, and neurons 2352 neurons.

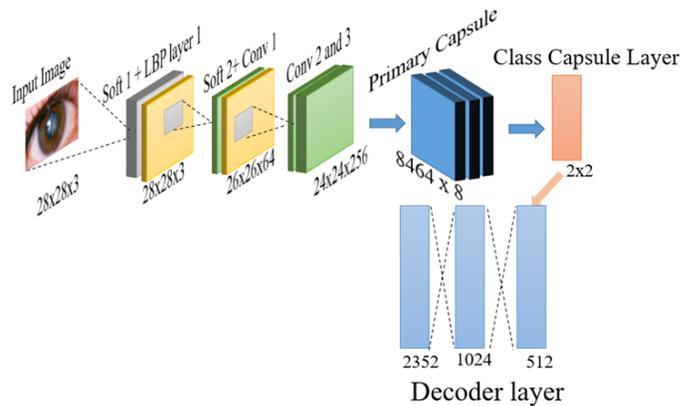


Fig. 3. The proposed model. Soft layer refers to the Scaled processing algorithm layer, Lbp refers to the local binary pattern layer and Conv refers to the convolutional layer.

F. Dataset Description and Preprocessing

The proposed model was evaluated on two open datasets downloaded from Kaggle^{1, 2}. The first dataset consists of 134 glaucoma images and 386 non-glaucoma compromised retinal images. The second dataset comprises 306 cataract images and 306 non-cataract compromised front eye images. These two datasets are combined and trained with the proposed model. Fig. 4 shows sample images from the dataset.

¹Sabari (2024). Cataract Dataset [online]. Website: <https://www.kaggle.com/datasets/sabari50312/fundus-pytorch> [accessed 5th April 2024]

²Siddharth P, Amit H, Dhaivat J (2024). Glaucoma Dataset [online]. Website: <https://www.kaggle.com/datasets/nandanp6/ataract-image-dataset> [accessed 5th April 2024]

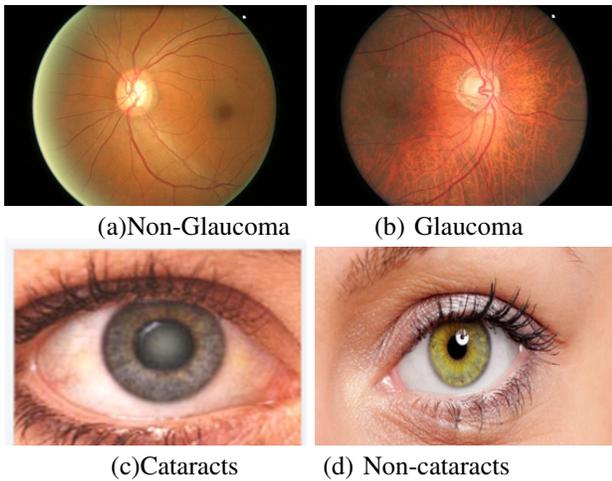


Fig. 4. Samples images from the combined dataset. (a) non-glaucoma compromised eye, (b) glaucoma compromised eye, (c) cataracts compromised eye, and (d) non-cataracts compromised eye.

G. Experimental Setting

The system used for the experiments has an NVIDIA GeForce 1060 with 8 Giga byte Random Access Memory. The batch size, learning rate, and learning decay rate were set to 100, 0.001, and 0.9, respectively. All codes were written using Keras with TensorFlow backend. The code at³ was adopted and modified to achieve the objectives of this paper. We adopted the margin loss function proposed by Sabour et al.[9] and trained their model which we refer to as a baseline model in the results and discussion section.

IV. RESULTS AND DISCUSSION

This section presents the results of our experimental evaluation, showcasing the performance of our proposed approach through various metrics and visualizations. We begin by analyzing the experimental curves, which illustrate the convergence and accuracy of our model. Next, we delve into the confusion matrix, which provides insights into the classification performance and error patterns. Ablation studies are then presented to dissect the contributions of individual components and hyperparameters to our model's success. Furthermore, we explore visual interpretability techniques to gain a deeper understanding of our model's decision-making processes. Finally, we compare our approach with state-of-the-art methods, demonstrating its competitive advantages and potential for future improvements.

A. Experimental Curves and Confusion Matrix Analysis

Fig. 5 presents the accuracy and loss curves for the proposed and baseline models trained on the combined, glaucoma-only, and cataract-only datasets. The proposed model attained 94.32%, 95.23%, and 96.87% on the combined, glaucoma-only, and cataract-only datasets, respectively. The baseline model attained 83.40%, 90.89%, and 92.00% on the combined, glaucoma-only, and cataract-only datasets, respectively. The

spikes in the curves of the proposed model are less and not intense compared to the curves (especially the training accuracy curves) of the baseline models. This shows the robustness of the proposed model in capturing the complex pattern in the data. Fig. 6 presents the confusion matrices for both models. Considering Tables 1 and 2, the proposed model had high precision and high sensitivity for all the classes while the baseline model had high sensitivity for all classes but lower precision for the cataract-positive class, lower specificity, and lower accuracy per class for all classes. The proposed model outperforms the baseline model in terms of precision, specificity, and accuracy per class for all classes. The proposed model shows slightly lower sensitivity for the glaucoma-positive class; however, it shows significant improvement in the precision of the glaucoma-positive class. Though the baseline model attained slightly higher sensitivity for the glaucoma-positive class, this is offset by its lower precision and specificity. Considering these analyses, the proposed model demonstrates better performance than the baseline model with improvements in precision, specificity, and accuracy per class for all the classes. In a layman's terms, the analysis of these metrics shows that, the proposed model is precise (has fewer false positives), more accurate (has fewer false negatives) and it's better at identifying specific eye problems (thus, glaucoma and cataracts). In a layman's terms, the analysis of the metrics of the baseline models shows an intelligent algorithm that can see clearly but not perfectly (more wrong prediction) while the proposed model is like a super-powerful microscope that helps see clearly and accurately (fewer wrong predictions).

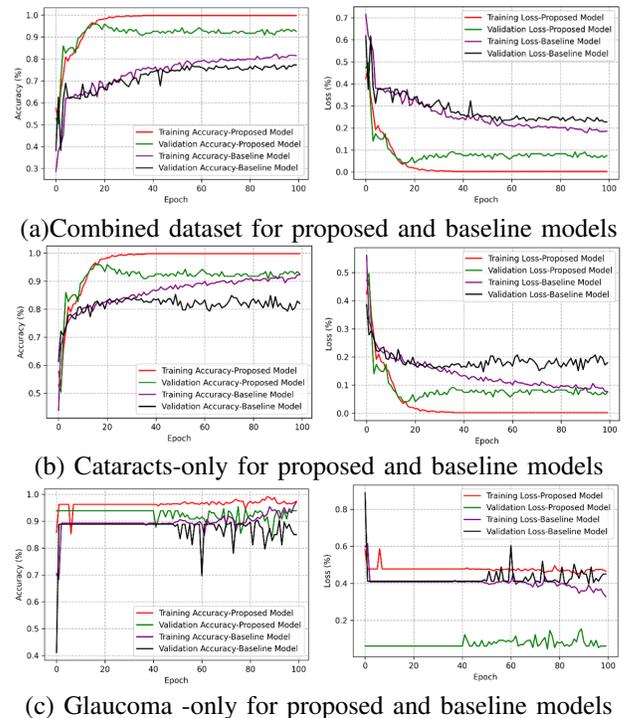


Fig. 5. Accuracy and loss curves for the proposed and baseline models.

B. Ablation Study

In this section, we conduct a systematic ablation study to dissect the contributions of each component in our proposed

³Xigenfuo (2018). Capsule Network Code [online]. Website: <https://github.com/XifengGuo/CapsNet-Keras> [accessed 25th February 2024]

TABLE I. ANALYZING THE CONFUSION MATRIX VALUES OF THE PROPOSED MODEL

Class	TP	FP	FN	TN	Precision	Sensitivity	Specificity	Accuracy per class
Glaucoma- Positive	23	1	4	201	0.958	0.859	0.995	0.978
Glaucoma-Negative	75	1	3	150	0.987	0.962	0.993	0.983
Cataracts-Positive	58	3	4	164	0.951	0.935	0.982	0.969
Cataract-Negative	60	6	2	221	0.909	0.968	0.974	1.00

TABLE II. ANALYZING THE CONFUSION MATRIX VALUES OF THE BASELINE MODEL

Class	TP	FP	FN	TN	Precision	Sensitivity	Specificity	Accuracy per class
Glaucoma-Positive	25	13	2	189	0.658	0.936	0.936	0.934
Glaucoma-Negative	59	5	19	146	0.921	0.967	0.967	0.895
Cataract-Positive	54	12	8	155	0.818	0.928	0.928	0.913
Cataract-Negative	53	8	9	159	0.869	0.952	0.952	0.926

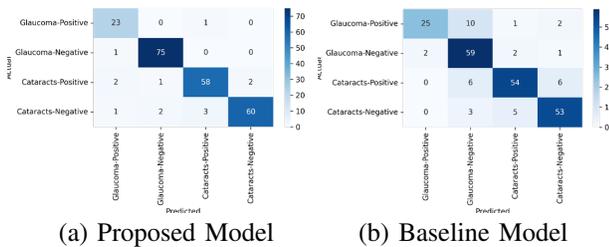


Fig. 6. Confusion matrix of the proposed and baseline models for the combined dataset.

Capsule Network algorithm. By meticulously examining the effects of removing or modifying individual components, we aim to uncover the crucial elements that drive the performance of our model. This rigorous analysis allows us to:

- Validate the design choices made in our algorithm,
- Identify the most critical components responsible for its effectiveness,
- Provide insights into the robustness and generalization capabilities of our approach,
- Offer a comprehensive understanding of the interactions between different components,

Through this ablation study, we demonstrate the importance of our novel feature enhancement algorithm (the scaled processing algorithm), and the incorporation of the local binary pattern layer, providing a deeper understanding of our Capsule Network algorithm’s inner workings and its ability to accurately diagnose glaucoma and cataract from fundus and non-fundus images. The values reported in Table III represent the validation accuracies attained as a result of the removal of a layer. The following were observed after the ablation study experiments:

- Scaled processing layers are crucial: Removing either Scaled processing layers (Soft 1 or Soft 2) or both leads to a significant drop in validation accuracy across all datasets. This indicates that the softmax activation function plays a vital role in enhancing contrast and improving feature extraction.

- Conv1 and LBP layer 1 are important for cataracts dataset: Removing Conv1 and LBP layer 1 together or separately affects the cataracts dataset more significantly than the other datasets. This suggests that these layers are essential for extracting features relevant to cataract detection.
- Conv2 and Conv3 are important for glaucoma dataset: Removing Conv2 and Conv3 together affects the glaucoma dataset more significantly than the other datasets. This indicates that these layers are crucial for extracting features relevant to glaucoma detection.
- LBP layer 1 is important for generalization: Removing LBP layer 1 affects all datasets, indicating its importance in improving the model’s generalization capabilities.
- Combining layer removals has a compounding effect: Removing multiple layers together (e.g. Soft 1 and Conv1, or Soft 1, 2, and LBP layer 1) leads to a more significant drop in validation accuracy than removing individual layers. This suggests that the layers work together to contribute to the model’s performance.

C. Visual Analysis

Visual examination of the feature maps (see Fig. 8) reveals that the scaled processing algorithm and local binary pattern (LBP) layer are the primary contributors to the clear and informative feature extraction, evident from the distinct shadow of the input images in the feature maps. The proposed model’s effective incorporation of these components enables robust feature extraction, and relevant feature selection, surpassing the baseline models’ feature maps, which exhibit blackout regions and reduced clarity. The reconstructed images (see Fig. 7) generated by the proposed model are clear and exhibit high certainty in class membership, unlike the baseline models. This demonstrates the proposed model’s ability to accurately capture and represent the underlying patterns in the data. The proposed model produces distinct clusters, albeit not compact (see Fig. 9a), indicating effective separation of classes. In contrast, the baseline model produces compact clusters (see Fig. 9b) but with significant cluster contamination, where members of one cluster are incorrectly assigned to other clusters. The proposed model prioritizes cluster distinctness

TABLE III. ABLATION STUDY RESULTS OF THE PROPOSED MODEL EXPERIMENTED WITH THE THREE DATASETS. “*” REPRESENTS THE REMOVAL OF A LAYER

Layer	Combined Dataset (%)	Glaucoma Dataset (%)	Cataracts Dataset (%)
*Soft 1	91.50	93.45	93.77
*Soft 2	90.86	92.57	92.49
*Soft 1 and 2	87.68	89.40	86.34
*Soft 1 and Conv1	89.47	92.56	93.68
*Soft 1, 2 and LBP layer 1	85.78	88.56	89.82
*Conv 2 and 3	92.34	93.71	95.67
*LBP layer 1, Soft 2 and Conv 1	91.39	90.23	92.43

and accuracy over compactness, particularly in applications where misclassification can have significant consequences. The proposed model’s ability to produce distinct clusters and accurately reconstruct images demonstrates its effectiveness in identifying and separating underlying patterns in the data.

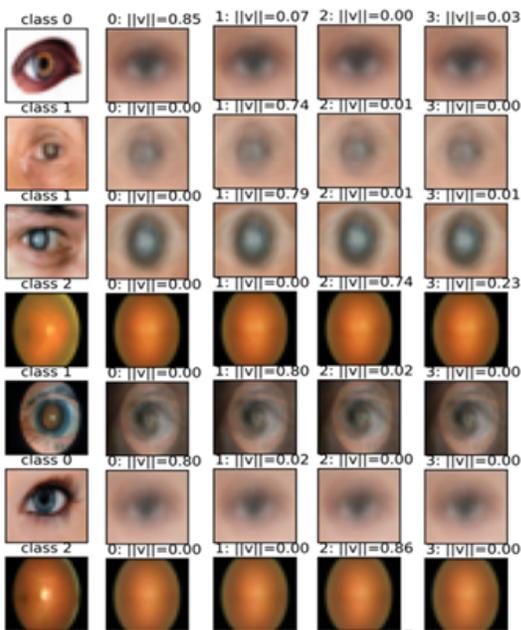


Fig. 7. Reconstructed images alongside their predicted classes for the (a) proposed model.

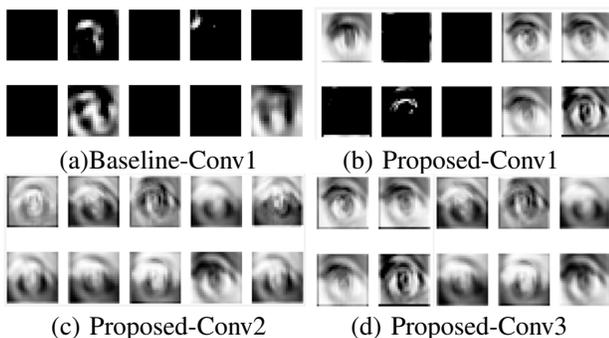


Fig. 8. Feature maps for the (a) Conv layer for baseline model, (b) Conv layer 1 for proposed model, (c) Conv layer 2 for proposed model and (d) Conv layer 3 for proposed model.

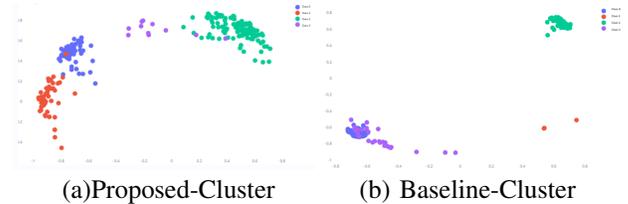


Fig. 9. Clusters of the decoder layer for (a) the proposed model and (b) the baseline model.

D. Comparison with other Works

Our proposed model achieves state-of-the-art performance on the combined dataset, outperforming all existing works. Notably, it surpasses the recent works of [12], [13] by a significant margin of 0.76% and 1.25%, respectively, on the cataracts dataset. Moreover, our model demonstrates a substantial improvement of 3.92% and 1.77% over the best-performing models of [1] and [14], respectively, on the glaucoma dataset. These results underscore the effectiveness of our proposed model in diagnosing both glaucoma and cataracts, showcasing its potential to improve patient outcomes in clinical settings. The superior performance of our model can be attributed to the novel feature enhancement algorithm and the incorporation of the local binary pattern layer, which enable more accurate feature extraction and improved robustness to variations in fundus and non-fundus images.

V. CONCLUSION

This paper presents a modified Capsule Network algorithm with a novel feature enhancement technique, termed the scaled processing algorithm, to diagnose glaucoma and cataract from fundus images. The proposed model addresses the limitations of existing methods by introducing a robust and efficient approach to feature extraction and illumination variation handling. The incorporation of a local binary pattern layer ensures reliability in diagnosis, while the scaled processing algorithm enhances feature maps, leading to improved recognition accuracy. The proposed model demonstrates comparable performance to state-of-the-art methods, achieving high accuracy on combined, cataract-compromised, and glaucoma-compromised eye datasets. This work contributes significantly to the field of medical image analysis, offering a promising tool for ophthalmologists and glaucoma specialists to accurately diagnose and manage glaucoma and cataract-compromised eyes, ultimately improving patient outcomes. The proposed approach has the potential to be extended to other medical image analysis

TABLE IV. COMPARISON OF THE PERFORMANCE OF THE PROPOSED MODEL TO OTHER WORK IN THE LITERATURE

Work	Combined Dataset (%)	Glaucoma Dataset (%)	Cataracts Dataset (%)
Baseline	83.40	90.89	92.00
De Santos et. al[1]	*	90.90	*
Sánchez-Morales et al.[15]	*	90.42	*
Liao et al. [16]	*	88.00	*
Lima et al.[17]	*	91.00	*
Chaudhary et al.[14]	*	91.10	*
de Sales et al.[18]	*	83.23	*
Fan et al.[19]	*	91.00	*
Jun et al.[20]	*	*	68.36
Wang et al.[13]	*	*	95.06
Wang et al.[12]	*	*	94.12
Proposed Model	94.32	95.23	96.87

applications, further highlighting its significance and impact. The proposed model currently lacks the capability to output uncertainties, which we aim to address in our future work.

REFERENCES

- [1] P. R. S. dos Santos, V. de Carvalho Brito, A. O. de Carvalho Filho, F. H. D. de Araújo, A. L. Rabêlo R de, M. J. Mathew, "A Capsule Network-based for identification of Glaucoma in retinal images," 2020 IEEE Symposium on Computers and Communications (ISCC), 2020, 1–6.
- [2] S. Cao, Y. Yao, G. An, "E2-Capsule Neural Networks for Facial Expression Recognition Using AU-Aware Attention," IET Image Processing, 2019, 14(11) 2417-2424.
- [3] C. Oguz, T. Aydin, M. Yaganoglu, "A CNN-based hybrid model to detect glaucoma disease," Multimedia Tools and Applications, 2024, 83(6), 17921–17939.
- [4] V. K. Velpula, L. D. Sharma, "Multi-stage glaucoma classification using pre-trained convolutional neural networks and voting-based classifier fusion," Frontiers in Physiology, 2023, 14, 1175881.
- [5] A. Shoukat, S. Akbar, S. A. Hassan, S. Iqbal, A. Mehmood, Q. M. Ilyas, "Automatic diagnosis of glaucoma from retinal images using deep learning approach," Diagnostics, 2023, 13(10), 1738.
- [6] D. J. Gaddipati, A. Desai, J. Sivaswamy, K. A. Vermeer, "Glaucoma assessment from oct images using capsule network," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, 5581–5584.
- [7] M. A. Ayidzoe, Y. Yongbin, P. K. Mensah, C. Jingye, K. Adu, T. Nyima, "Feature amplification capsule network for complex images," Journal of Intelligent & Fuzzy Systems, 2021, <https://doi.org/10.3233/JIFS-202080>.
- [8] G. Hinton, A. Krizhevsky, S. D. Wang, "Transforming Autoencoders," International Conference on Artificial Neural Networks Springer, 2011, 44–51. https://doi.org/10.1007/978-3-642-21735-7_6
- [9] S. Sabour, N. Frosst, G. Hinton, "Dynamic Routing Between Capsules," Advances in Neural Information Processing Systems, 2017, 3856–3866.
- [10] T. Ojala, M. Pietikäinen, T. Mäenpää, "Multiresolution Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 1–35.
- [11] K-N. Nguyen-Thi, H. Che-Ngoc, A-T. Pham-Chau, "An efficient image contrast enhancement method using sigmoid function and differential evolution," Journal of Advanced Engineering and Computation, 2020, 4(3), 162–172.
- [12] Y. Wang, C. Tang, J. Wang, Y. Sang, J. Lv, "Cataract detection based on ocular B-ultrasound images by collaborative monitoring deep learning," Knowledge-Based Systems, 2021, 231, 107442.
- [13] T. Wang, J. Xia, R. Li, R. Wang, N. Stanojic et al, "Intelligent cataract surgery supervision and evaluation via deep learning," International Journal of Surgery 2023, 104, 106740.
- [14] P. K. Chaudhary, R. B. Pachori, "Automatic diagnosis of glaucoma using two-dimensional Fourier-Bessel series expansion based empirical wavelet transform," Biomedical Signal Processing and Control, 2021, 64, 102237.
- [15] A. Sánchez-Morales, J. Morales-Sánchez, O. Kovalyk, R. Verdú-Monedero, J. L. Sancho-Gómez, "Improving Glaucoma Diagnosis Assembling Deep Networks and Voting Schemes," Diagnostics 2022, 12(6), <https://doi.org/10.3390/diagnostics12061382>.
- [16] W. Liao, B. Zou, R. Zhao, Y. Chen, Z. He, M. Zhou, "Clinical interpretable deep learning model for glaucoma diagnosis," IEEE Journal of Biomedical and Health Informatics, 2019, 24(5), 1405–1412.
- [17] A. Lima, L. B. Maia, P. T. C. dos Santos, G. B. Junior, J. D. S. de Almeida, A. C. de Paiva, "Evolving convolutional neural networks for glaucoma diagnosis," Anais Do XVIII Simpósio Brasileiro de Computação Aplicada à Saúde, 2018.
- [18] N. R. de Sales Carvalho, C. da. M. Rodrigues, A. O. de Carvalho Filho, M. J. Mathew, "Automatic method for glaucoma diagnosis using a three-dimensional convoluted neural network," Neurocomputing, 2021, 438, 72–83.
- [19] R. Fan, K. Alipour, C. Bowd, M. Christopher, N. Brye, J. A. Proudfoot, M. H. Goldbaum, A. Belghith, C. A. Girkin, M. A. Fazio, "Detecting glaucoma from fundus photographs using deep learning without convolutions: transformer for improved generalization," Ophthalmology Science, 2023, 3(1), 100233.
- [20] T. J. Jun, Y. Eom, C. Kim, D. Kim, "Tournament based ranking CNN for the cataract grading," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, 1630–1636.