Enhanced Glaucoma Detection: Ensemble Knowledge Distillation Approach

Mohammed Arif Uddin
2012967042
mohammed.uddin05@northsouth.edu

Al Ibne Siam
2012072042
ibne.siam@northsouth.edu

Zinnat Fowzia Ria
1931343042
zinnat.ria@northsouth.edu

Md Hasibur Rahman
2012558042
md.rahman@northsouth.edu

***Abstract*—Detecting physiological changes within the human body poses a significant challenge in biomedical engineering. Currently, these irregularities necessitate manual grading, a laborious and time-intensive process due to the intricate methods involved in their assessment. Glaucoma is a major reason for permanent vision loss worldwide, caused by high pressure in the eyes. Detecting it accurately and quickly is crucial to prevent vision problems. But finding glaucoma manually is tough and needs a lot of skill and experience. To enable early-stage illness identification, there is a growing focus on computer-assisted diagnostics, aiming to develop a disease detection system. In this paper, we combine the strength of ensemble methods with the effectiveness of knowledge distillation. The main concept involves training top-performing models, distilling their extensive knowledge, and then merging these refined models into a unified one.**

The experiment results reveal our multi-faceted approach in glaucoma detection, relying on two primary techniques. Initially, knowledge distillation from a teacher model was used to create two distinct student models, later combined into a single unified model. The teacher model achieved 80% accuracy and a 79% macro F1 score. Student 1, based on ResNet18, initially achieved 68% accuracy and a 65% macro F1 score without distillation. However, after distillation, significant improvements were seen, reaching 71% accuracy and a 69% macro F1 score. Meanwhile, Student 2, leveraging EfficientNetB0, attained 63% accuracy and a 62% macro F1 score initially, which notably increased to 76% accuracy and a 70% macro F1 score post-distillation. Following the ensemble fusion of Student 1 and Student 2, a remarkable enhancement emerged, with the accuracy rising to 79%, accompanied by a 78% Weighted F1 Score and an impressive macro F1 score of 74%, closing the gap with the teacher model's accuracy. Most significantly, the ensembled student model only has a parameter count of 16.9 million compared to the 136 million parameters of the teacher VGG16 model, thus achieving an 8 times reduction in parameters while keeping relatively same performance.

Keywords: Knowledge Distillation, Ensemble, Glaucoma, RIM-ONE, Fundus.

I. Introduction

Glaucoma stands as a primary contributor to permanent vision loss in our aging population, expected to affect 112 million individuals by 2040 [1]. This long-term nerve condition causes damage to the optic nerve fibers, resulting in noticeable alterations within and around the optic disc, ultimately resulting in loss of functional vision. Glaucoma involves distinct alterations in the optic nerve head (ONH), commonly referred to as the optic disc [2]. At present, diagnosing and monitoring glaucoma involves comprehensive eye exams and extensive data collection, posing challenges in interpretation. Moreover, distinguishing between normal ocular characteristics and early glaucoma features often overlaps. Hence, there's a growing interest in developing additional methods, like artificial intelligence (AI) [3] systems, to aid in distinguishing genuine pathology from normal variations and actual progression from test-related differences.

With this objective, our paper introduces a framework utilizing knowledge distillation and ensemble methods to elevate the effectiveness and efficiency of glaucoma detection models. This study aims to enhance classification accuracy by automatically categorizing normal and glaucomatous eye photos using fundus image. To achieve higher precision, it's crucial to explore and select specific and sensitive criteria for categorizing glaucoma images, surpassing the current standard methods.

Contributions: Our main contributions can be summarized as follows:

1) To our understanding, our framework represents the initial fusion of knowledge distillation and ensemble techniques aimed at enhancing the efficiency and effectiveness of Large Models for practical glaucoma detection.

2) We distill the knowledge from a large teacher model into

multiple smaller models and finally ensemble them to achieve results comparable to teacher but with only an eight of its parameters,thus reducing complexity significantly.
3) We show a comparison between the superior performance of multiple distilled models and their ensemble compared to baseline models,thus proving knowledge distillation and ensemble as robust and efficient methods.

II. Relevant Work

In this section, we examine key literature across four specific areas: Glaucoma detection on RIM-ONE dataset, Glaucoma detection using Machine Learning, knowledge distillation and related methods, and ensemble methods for glaucoma detection.

Glaucoma Detection on RIM-ONE Dataset. In this [4] paper, they conducted training on five convolutional neural network (CNN) models—standard CNN, VGG19, ResNet50, GoogleNet, and DENet—utilizing the RIM-ONE dataset. Among these models, VGG19 exhibited superior performance in glaucoma detection, achieving an AUC (Area Under the Curve) of 0.94, a sensitivity of 87.0%, and a specificity of 89.0%. Using the ResNet-50 architecture, a study achieved excellent results for detecting glaucoma on the G1020, RIM-ONE, ORIGA, and DRISHTI-GS datasets. They obtained a detection accuracy of 98.48%, a sensitivity of 99.30%, a specificity of 96.52%, an AUC of 97% [5].

Eyecare using Machine Learning. Research in the application of predictive analytics in eyecare encompasses the following areas: forecasting post-operative surgical outcomes [6] [7] [8], disease prediction(Glaucoma [9] [10], AMD [11] [12]), segmentation of eye parts and anomalies (blood vessels [13]).

Knowledge Distillation and Related Methods. A study achieved accuracy of 99% using Le-Net for input image validation. Considering the application of brightest spot algorithm, an accuracy of 98.67% is achieved for ROI extraction [14]. Another study uses Deep Learning for detecting Glaucoma. Their results showed that the proposed method can identify glaucoma from eye fundus images with an accuracy of 90.0% [15]. A study proposes ensemble technique for detecting glaucoma. They used different dataset and accuracy rates of 95.63%, 98.67%, 95.64%, and 88.96% were achieved using the DRIONS-DB, HRF, DRISHTI-GS, and combined data sets, respectively [16].

III. Methods

3.1. Overview

Knowledge distillation is a process in machine learning where a smaller, more lightweight model learns from a larger, more complex one. Through distillation, the smaller model aims to replicate the larger model's performance by capturing its essential knowledge, often leading to improved efficiency and faster inference times. Ensemble methods stand as a cornerstone in contemporary machine learning, enabling the fusion of varied models to transcend individual constraints and achieve unparalleled levels of model performance. Knowledge distillation holds significance in the medical sector as it facilitates the deployment of computationally lighter models without compromising performance. In healthcare applications where resources and computational power may be limited, distillation enables the creation of efficient models that can aid in faster diagnoses, treatment planning, and real-time decision-making, enhancing the scalability and practicality of machine learning solutions within medical settings. This approach not only preserves the performance capabilities of a larger model (such as an ensemble) but also notably minimizes the computational resources required. In this study, our focus centers on practical applicability by employing a methodology that involves distilling knowledge from a VGG16 model. Utilizing VGG16 as the teacher model, we transferred its knowledge to create multiple student models. Specifically, Student model 1 adopts ResNet18 architecture, while Student model 2 is based on EfficientNetB0. Subsequently, we combined Student model 1 and Student model 2 into an ensemble, significantly reducing the parameters by eight fold compared to the original model.

3.2. VGG16

[17] introduced VGG16. The VGG16 architecture, comprising 16 layers, is structured with two main components. Initially, a 224×224 RGB image undergoes processing through a sequence of convolution layers, maintaining its size at 224×224 RGB. The initial layers consist of 3×3 filters with 64 channels before transitioning to subsequent layers. Following a max-pool layer with a stride of (2, 2), two layers of 3×3 filters

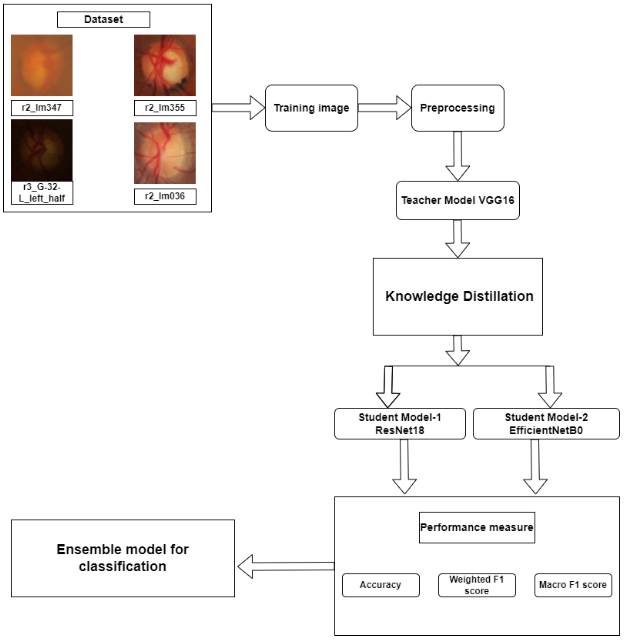

Fig. 1. Comprehensive overview of the research methodology

and a convolution layer employing 256 filters are applied. This sequence is then repeated twice with sets of three convolution layers and a max-pool layer. Notably, the VGG16 architecture opts for 3×3 filter sizes in lieu of larger sizes like 7×7 or 11×11, employing 1 pixel padding after each convolution layer to preserve spatial information in the image.

The network layout of VGGNet-16, illustrated in Figure 2, features a cascade of convolution layers followed by three fully connected layers, succeeded by another series of convolution layers. The first and second fully connected layers contain 4,096 channels each, with the first layer having the highest number of channels. The third layer, linked to the SoftMax layer, comprises 1,000 channels.

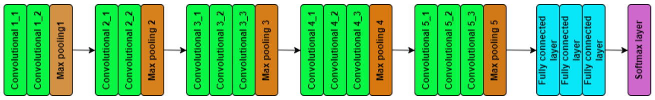

Fig. 2. ConvNet architecture: VGG-16.

3.3. *EfficientNetB0*

[18] introduced EfficientNetB0. EfficientNet represents a convolutional neural network architecture and scaling approach that uniformly adjusts all aspects of depth, width, and resolution by utilizing a compound coefficient. They utilized neural architecture search to create a new baseline network and subsequently scaled it up, resulting in a range of models termed EfficientNets. These models showcased significantly improved accuracy and efficiency compared to previous ConvNets. Specifically, their EfficientNet-B7 achieved a top-1 accuracy of 84.3% on ImageNet, while being 8.4 times smaller and 6.1 times faster during inference than the leading existing ConvNet. Furthermore, their EfficientNets demonstrated exceptional transferability, achieving state-of-the-art accuracy on CIFAR-100 (91.7%), Flowers (98.8%), and three other transfer learning datasets, all with substantially fewer parameters.

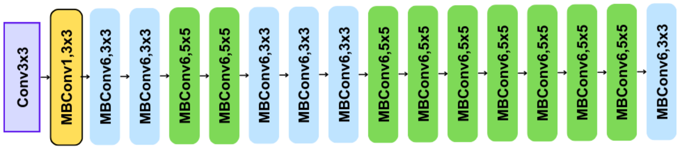

Fig. 3. ConvNet architecture: EfficientNetB0.

3.4. *ResNet18*

[19] introduced ResNet18. They introduce a residual learning framework aimed at simplifying the training process for networks significantly deeper than those previously employed. Their approach involves explicitly restructuring the layers as learning residual functions in relation to the layer inputs, rather than learning functions without references. Through extensive empirical evidence, they demonstrate that these residual networks are more amenable to optimization and can achieve increased accuracy with substantially greater depth. Assessing residual nets with a depth of up to 152 layers on the ImageNet dataset—eight times deeper than VGG nets but with lower complexity—they attain a 3.57% error rate on the ImageNet test set. This accomplishment secured the 1st place in the ILSVRC 2015 classification task. Additionally, they provide analyses on CIFAR-10, exploring models with 100 and 1000 layers.

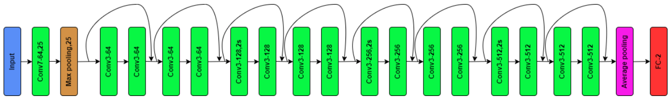

Fig. 4. ConvNet architecture: Resnet18.

3.5. *The Distillation Network*

In the process of knowledge distillation, we leverage the VGG16 Network as teacher model. Then we distilled the knowledge into a multiple student model. Our first student model is RestNet18 and second student model is Effiecient-NetB0. The relationship between the student and teacher models is determined by comparing their outputs for the identical input. We incorporate a temperature parameter T to adjust the logits before applying the SoftMax function in both the student and teacher models. The softened probability distribution is computed as follows:

$$\sigma(z_x) = \frac{\exp(\frac{x_i}{\tau})}{\sum_j \exp(\frac{x_j}{\tau})}; i = 0, 1, \ldots, n \tag{1}$$

where zi represents the logits, while K signifies the number of classes. The temperature parameter T governs the softness of the probability distribution. When T equals 1, the SoftMax operates conventionally; however, for T greater than 1, it generates a softer distribution. This softening effect prompts the student model to prioritize understanding relationships among various classes instead of solely concentrating on the accurate classification. The loss function used in knowledge distillation typically combines the standard classification loss (such as cross-entropy with the true labels) and the distillation loss (for instance, KL divergence between the softened probabilities of the student and teacher models).

3.6. The Ensemble Network

The ensemble approach stands as a robust and efficient method that constructs multiple individual classification models, subsequently amalgamating them to enhance accuracy. We define ensemble model as a series of student model. ResNet18 and EffecientNetB0 are our student model and combine them to create our ensemble model.

IV. Experiments and Results

4.1. Dataset Details

The inception of the Retinal Image database for Optic Nerve Evaluation (RIM-ONE) dates back to 2011, followed by two subsequent iterations, solidifying its status as one of the most referenced public retinography databases for assessing glaucoma. Originally designed as a repository for optic disc segmentation reference images, its recent predominant use has shifted towards training and validating deep learning models. With the REFUGE challenge setting specific criteria for such data sets' validation, coupled with observed confusion and misuse across the three versions, a decision emerged to consolidate and revise them into a new iteration, coined RIM-ONE DL (RIM-ONE for Deep Learning). This paper details this image collection, comprising 313 retinographies from healthy individuals and 172 from glaucoma patients, all meticulously evaluated by two experts and accompanied by manual disc and cup segmentations. Additionally, it presents an evaluation framework utilizing various well-established convolutional neural network models [20].

4.2. Data Pre-processing

For glaucoma detection using the RIM-ONE dataset, initial data preprocessing involves loading the retinal images and standardizing them by resizing to a uniform dimension, converting to a consistent format and normalizing pixel values. These images are then labeled based on their respective classes ('glaucoma' or 'normal'), augmented for diversity using techniques like rotation and flipping, and split into training, validation, and test sets. Quality checks to remove corrupted or low-quality images ensure dataset integrity. Feature extraction methods like optic disc or cup segmentation applied, generating relevant inputs for models. This comprehensive preprocessing pipeline optimizes dataset readiness for glaucoma detection model training and evaluation.

4.3. Experimental Setup

4.3.1. Implementation Details Our experiments were run using a virtual GPU P100 and 13 GB of virtual memory and 32 GB of RAM.

4.3.2. Training and Evaluation Evaluation of each deep learning architecture's performance relies on metrics encompassing accuracy, specificity, precision, sensitivity, and the F1 score. These metrics are quantified by the following mathematical expressions: [utf8]inputenc amsmath

$$\begin{aligned} Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\ Sensitivity &= \frac{TP}{TP + FN} \\ Precision &= \frac{TP}{TP + FP} \\ Specifity &= \frac{TN}{TP + FP} \\ F1score &= \frac{2 * (sensitivity * precision)}{sensitivity + precision} \end{aligned} \tag{2}$$

where TP represents the number of correctly predicted positive cases, TN signifies the number of correctly predicted negative cases, FN indicates the instances incorrectly predicted as negative, and FP denotes the cases incorrectly predicted as positive.

4.4.Results

The experiment results outlined in Table 2 showcase our comprehensive approach in glaucoma detection, primarily employing two techniques. Initially, we distilled the knowledge from a teacher model into two distinct student models, subsequently amalgamating them into a unified model. The teacher model exhibited 80% accuracy and a 79% macro F1 score. Student 1, utilizing ResNet18, initially achieved 68% accuracy and a 65% macro F1 score without distillation. Remarkably, with distillation, significant improvements were evident, reaching 71% accuracy and a 69% macro F1 score for the student 1 network. Meanwhile, Student 2, leveraging EfficientNetB0, achieved 63% accuracy and a 62% macro F1 score without distillation, which notably surged to 76% accuracy and a 70% macro F1 score post-distillation. Following the ensemble of Student 1 and Student 2, a remarkable enhancement occurred, bringing the accuracy closer to that of the teacher model, achieving 79% accuracy, a 78% Weighted F1 Score, and a notable macro F1 score of 74%.

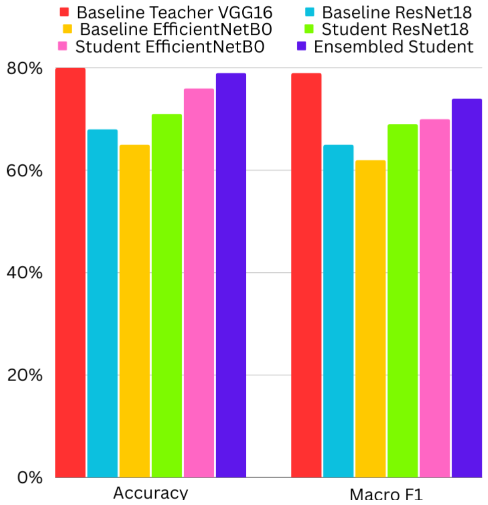

Fig. 5. Comparison of accuracy of the proposed model.

We furthered our glaucoma detection exploration by incorporating ResNet50, VGG16, InceptionNetV3, and DenseNet202 individually, then combining their strengths to forge an ensemble model. Throughout our experiment, we diligently compared various epoch sizes and ensemble combinations to scrutinize their impact. Our rigorous analysis revealed a peak performance, reaching 80% accuracy, a macro F1 score of 78%, and a weighted F1 score of 80%. Notably, employing ResNet50, VGG16, and InceptionV3 in tandem with an epoch size of 50 yielded the highest accuracy, surpassing other ensemble combinations—a significant discovery in our research.

Model	**Accuracy**	**Macro F1 Score**
Teacher VGG16	80%	79%
ResNet18	68%	65%
EfficientNetB0	63%	62%

TABLE I
EVALUATION MATRICES OF OUR PROPOSED METHODS WITHOUT DISTILLATION

Model	**Accuracy**	**Macro F1 Score**
Distilled Resnet18	71%	69%
Distilled EfficientNetB0	76%	70%
Ensembled Student	79%	74%

TABLE II
EVALUATION MATRICES OF OUR PROPOSED METHODS WITH DISTILLATION

Model	**Number of Parameters**
Teacher VGG16	136 million
Student1 ResNet18	11.6 million
Student2 EfficientNetB0	5.3 million
Ensembled Student	16.9 million

TABLE III
PARAMETERS OF TEACHER AND STUDENT MODELS

V. CONCLUSION AND FUTURE WORK

In conclusion, our study delved into glaucoma detection employing two distinct methodologies: knowledge distillation with two student models and an ensemble approach featuring multiple prominent architectures. Through knowledge distillation, we observed significant enhancements in student model performance, particularly after distilling insights from the teacher model. Furthermore, our ensemble strategy, combining ResNet50, VGG16, and InceptionV3, yielded the highest accuracy, showcasing the potential of this amalgamation for robust glaucoma detection. These findings underscore the efficacy of knowledge distillation in refining individual models and the power of ensemble techniques in harnessing diverse model strengths for enhanced accuracy in glaucoma detection.

Moving forward, we're eager to explore why distilled EfficientNetB0 performs better than distilled ResNet18 in our glaucoma detection study. We'll take a closer look at the different parts and ways these models work. By doing this, we hope to uncover what makes EfficientNetB0 stand out when we're teaching it using knowledge distillation. This investigation will help us understand the reasons behind their varying performance levels.

References

(1) Tham, Y.-C.; Li, X.; Wong, T. Y.; Quigley, H. A.; Aung, T.; Cheng, C.-Y. Global Prevalence of Glaucoma and Projections of Glaucoma Burden through 2040:A Systematic Review and Meta-Analysis, 2014.

(2) Group, O. H. T. S.; Group, E. G. P. S. Validated Prediction Model for the Development of Primary Open-Angle Glaucoma in Individuals with Ocular Hypertension, 2007.

(3) Greenfield, D. S.; Weinreb, R. N. Role of Optic Nerve Imaging in Glaucoma Clinical Practice and Clinical Trials, 2008.

(4) Gómez-Valverde, A. A. G. F. B. L. A. H. J. J.; Santos, A.; Sánchez, C. I.; Ledesma-Carbayo, M. J. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning, 2019.

(5) Shoukat, A.; Akbar, S.; Hassan, S. A.; Iqbal, S.; Mehmood, A.; Ilyas, Q. M. Automatic Diagnosis of Glaucoma from Retinal Images Using Deep Learning Approach, 2023.

(6) Rao, H. L.; Addepalli, U. K.; Yadav, R. K.; Choudhari, N. S.; Senthil, S.; Mandal, A. K.; Garudadri, C. S. Accuracy Of Ordinary Least Squares And Empirical Bayes Estimates Of Short Term Visual Field Progression Rates To Predict Long Term Outcomes In Glaucoma, 2012.

(7) Gupta, M.; Gupta, P.; Vaddavalli, P. K.; Fatima, A. Predicting Post-operative Visual Acuity for LASIK Surgeries, 2016.

(8) Dis, I. J. K. E. C. Predictors of Clinical Outcomes after Intrastromal Corneal Ring Segments Implantation, 2012.

(9) Bowd, C.; Weinreb, R. N.; Balasubramanian, M.; Lee, I.; Jang, G.; Yousefi, S.; Zangwill, L. M.; Medeiros, F. A.; Girkin, C. A.; Liebmann, J. M.; Goldbaum, M. H. Glaucomatous Patterns in Frequency Doubling Technology (FDT) Perimetry Data Identified by Unsupervised Machine Learning Classifiers, 2014.

(10) Fu, H.; Cheng, J.; Xu, Y.; Zhang, C.; Wong, D. W. K.; Liu, J.; Cao, X. Disc-Aware Ensemble Network for Glaucoma Screening From Fundus Image, 2018.

(11) Fraccaro, P.; Nicolo, M.; Bonetto, M.; Giacomini, M.; Weller, P.; Traverso, C. E.; Prosperi, M.; OSullivan, D. Combining macula clinical signs and patient characteristics for age-related macular degeneration diagnosis: a machine learning approach, 2015.

(12) L., C. S.; BS, M. D. M. B.; MD, A. Y. L.; MSCI Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images, 2017.

(13) Fraz, M. M.; Remagnino, P.; Hoppe, A. An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation, 2015.

(14) Shinde, R. Glaucoma detection in retinal fundus images using U-Net and supervised machine learning algorithms, 2021.

(15) Bragança, C. P.; Torres, J. M.; d. A. Soares, C. P. Detection of Glaucoma on Fundus Images Using Deep Learning on a New Image Set Obtained with a Smartphone and Handheld Ophthalmoscope, 2022.

(16) Joshi, S.; Partibane, B.; Hatamleh, W. A.; Tarazi, H.; Yadav, C. S.; Krah, D. Glaucoma Detection Using Image Processing and Supervised Learning for Classification, 2022.

(17) Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014.

(18) Tan, M.; Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, Long Beach, California, 2019.

(19) He, K.; Zhang, X.; Ren, S.; Sun, J. Very Deep Convolutional Networks for Large-Scale Image Recognition, Las Vegas, NV, USA, 2016.

(20) Diaz-Aleman, T.; Sigut, J.; Alayon, S.; Arnay, R.; Angel-Pereira, D. RIM-ONE DL: A Unified Retinal Image Database for Assessing Glaucoma Using Deep Learning, 2020.