Auxiliary-Domain Learning for a Functional Prediction of Glaucoma Progression

Sean Wu[1], Vahid Mohammadzadeh[2], Kiumars Edalati[2], Jack Martinyan[2], Arthur Martinyan[2], Joseph Caprioli[2], Kouros Nouri-Mahdavi[2], and Fabien Scalzo[1,3]

[1] Seaver College, Pepperdine University, Malibu, CA 90265
fabien.scalzo@pepperdine.edu
[2] Glaucoma Division, Stein Eye Institute, David Geffen School of Medicine, University of California, Los Angeles (UCLA), CA 90095
[3] Department of Computer Science, University of California, Los Angeles (UCLA), CA 90095

Abstract. An accurate and early prediction of a patient's glaucoma progression can give ophthalmologists insight on how to mitigate the ramifications of the disease before they experience irreversible visual field loss or blindness. Our paper introduces an auxiliary-domain learning framework that trains a convolutional neural network to predict glaucoma progression (main task) and utilizes auxiliary tasks during training, including prediction of the patient age, mean deviation, and optical coherence tomography (OCT) data to improve its accuracy on the main task. The modalities of optic disc photographs and OCT data are often not utilized jointly due to costly machinery. However, we exploit informative features in the OCT, age, and mean deviation data as our learning objective to alleviate the need to acquire the data in clinical deployment. We compared baseline models with no auxiliary outputs to the ones built using auxiliary tasks, and observed a 6.5% increase in Area Under the Receiver Operating Characteristic Curve (AUC-ROC) in the final auxiliary-domain model (91.3±2.6%) compared to the baseline (84.8±4.9%). This study demonstrates the utility of auxiliary tasks when training ophthalmological models by leveraging important patient data that is difficult to acquire during training, even when it is not available as part of the model's deployment in routine clinical care.

Keywords: Glaucoma Progression · Auxiliary Domain Learning · Siamese Neural Networks

1 Introduction

Glaucoma is an optical disease responsible for the degeneration of the optic nerve and a gradual visual field loss. It is one of the leading causes of blindness worldwide [1], affecting about 2.2 percent of the world's population [2]. While there is no cure of this disease, early diagnosis has demonstrated the ability to prevent further visual field loss through surgery and other treatments. Conventional

computer vision techniques have been utilized in recent studies to diagnose progression of the disease by using optic retinal images [3]. While these methods show proficiency in classifying glaucoma progression, training a multi-task neural network has the potential to greatly enlarge the generalizability of a neural network. In this paper we discuss the benefits of hard parameter sharing between the convolutional and fully connected layers when training a deep learning model to classify glaucoma progression.

Machine learning has made a significant contribution to ophthalmic research, assisting ophthalmologists to treat various optic ailments such as diabetic retinopathy [4], glaucoma screening through optic disc photographs [5], and also volumetric deep learning algorithms to analyze optical coherence tomography scans [6]. In previous studies, visual field and clinical data such as mean deviation have been used to assess glaucoma progression [7]. Optic disc photographs have also shown promising prevalence in predicting symptoms of glaucoma [8]. In this paper, we constructed a deep learning model to utilize both optic disc photographs, clinical data, and OCT data as a multi-output neural network to enhance the predictability of glaucoma progression. A diagram of our methods is depicted in Figure 1 where we use additional data as model outputs to facilitate the training process.

The addition of multi-modal features to a deep learning model has been shown to positively impact the network's performance depending on how predictive that feature is to the particular task [9]. For glaucoma progression, the mean deviation index, optical coherence tomography (OCT) clock-hour acquisitions, and global eye thickness (computed from OCT) hold crucial information to whether the patient's glaucoma will progress or not. However, acquiring these additional inputs can be costly due to requiring an OCT machine and additional visual field tests. To address this challenge, we trained a model to learn these external features through a multi-output convolutional neural network to be able to utilize the important characteristics of those features in relation to glaucoma progression prediction but without explicitly needing them as an input (Figure 1). In this paper, we construct a model that captures multi-modal features in the training process and can be deployed on unseen data without the need of these additional, potentially costly and unavailable, modalities.

2 Methods

Our model consists of three parallel Siamese Neural Networks[10], where each encoding of the optic disc photographs "share" the exact copy of weights with one another. Additionally, we explore a composite loss function to weigh and propagate our multi-task losses through the model.

2.1 Data Acquisition

The dataset used in our experiments consists of 1,706 eyes acquired at the University of California, Los Angeles (UCLA) Jules Stein eye institute. Each sample

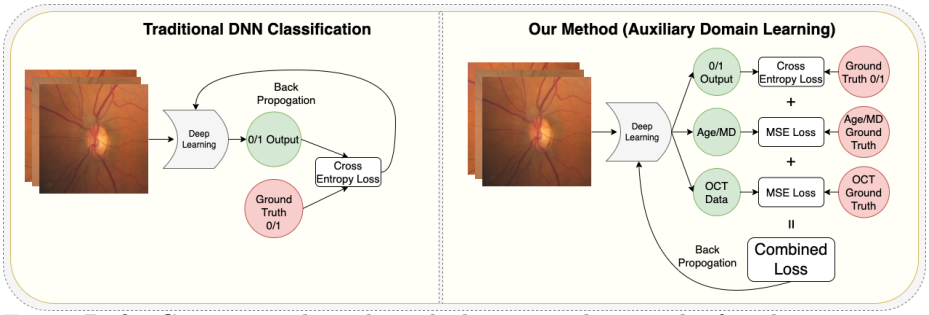

Fig. 1: **Left:** Conventional single task deep neural networks for glaucoma progression prediction with optic disc photographs. **Right:** Representation of our auxiliary-domain neural network to improve the prediction of glaucoma progression.

has a minimum of one optic disc photograph and more than five visual field exams with at least 3 years of follow-up data. Each eye's mean deviation index was calculated with a linear regression model, and the visual field progression was determined as a statistically significant negative slope ($p<0.05$) for at least two consecutive visits and at the final visit of the sequence. The three optic disc photographs that are inputted into our model were acquired at different visit dates, and the third visit photograph was taken at least one year before the diagnosed progression date. Due to the prolonged time between each visit, different devices were used to obtain the optic disc photographs. The Zeiss 450 camera (Carl Zeiss Meditec, Dublin, CA) and the FF 450 plus Fundus Camera with VISUPAC Digital Imaging System (Carl Zeiss Meditec, Dublin, CA) were the devices used to take the optic disc photographs. Each eye also has a corresponding retinal nerve fiber layer OCT data taken by the Cirrus SD-OCT machine that consists of thickness measurements for 12 clock-hour angles and an additional global thickness value that is the average of the clock-hour values. The use of this data was authorized by UCLA's Institutional Board Review.

2.2 Baseline Deep Learning Model

To train our model, we used the shallowest network in the residual neural network family ResNet18[11], pre-trained on Imagenet [12] as the backbone for our Siamese Neural Network (SNN). Our SNN consists of three identical networks, where each encoding of the optic disc photographs "share" their weights with one another. The 512 features that are extracted from each instance of the ResNet18 encoder are then concatenated with one another, creating a vector of 1536 elements that can be fully connected to the binary glaucoma progression neuron.

The 1706 eyes were first randomly shuffled and partitioned into an approximate [80%/10%/10%] training/validation/testing split. To standardize the data, the left eye images (OS) were mirrored along the horizontal axis to represent a

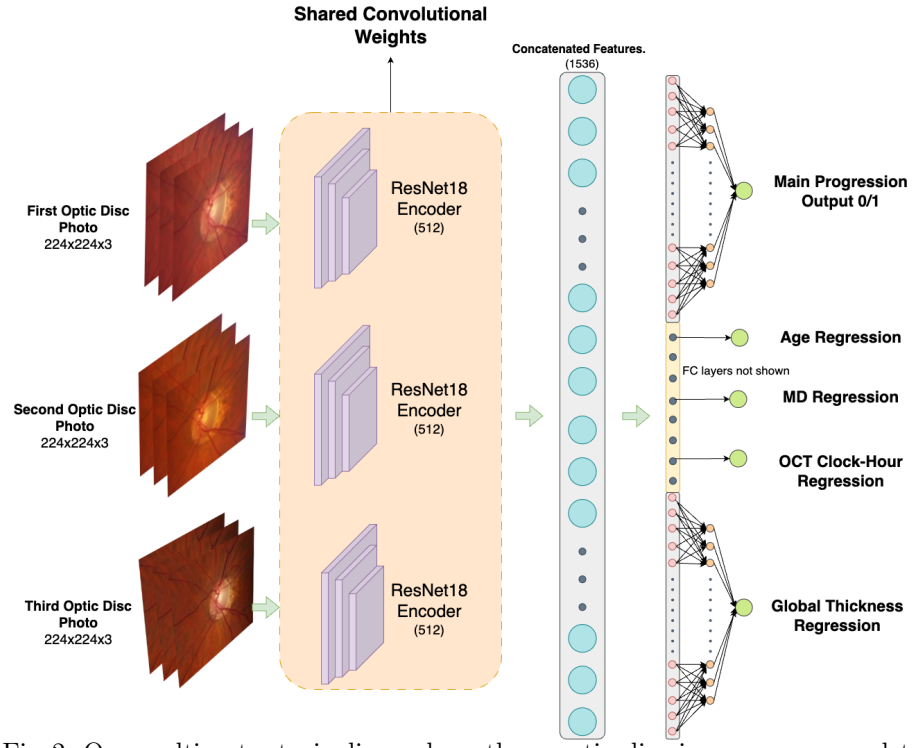

Fig. 2: Our multi-output pipeline, where three optic disc images are passed to a Siamese Neural Network with shared weights, then fed to their task-specific fully connected layers.

right eye (OD). This mirroring was done to reduce variance and variability in the data, helping guide the model for easier learning. Additionally, we applied data augmentation methods of shifting, scaling, and rotating geometric transformations with parameters of 0.0625, 0.2, and 90 degrees, respectively, to ensure that our model is invariant to any spatial changes and can generalize well on a broader population. In training, we set the batch size to 16 and the learning rate to 1×10^{-4}, and ran each network for approximately 100 epochs while saving the highest validation area under the receiver operating characteristic curve.

2.3 Hard Parameter Sharing and Combined Loss function

To ensure that our model undergoes multi-domain learning, we utilized a common hard parameter sharing technique, where the initial network layers are seen as a common ground between tasks[13]. Figure 2 illustrates our method of hard parameter sharing, where the ResNet18 encoder weights are shared in a SNN amongst the three optic disc photographs. If a progressing sequence did not meet the criteria for three optic disc photographs, a padded image of numerical

constants was used as a replacement. This is a common practice for temporal data such as glaucoma progression, where the number of visits for each patient is variant. To ensure consistent input shapes and prevent any input tensor mismatch errors during the forward pass to the model, we utilized blank tensors as placeholder inputs for these eyes. Sequence padding is a common approach for problems consisting of temporal data, such as glaucoma progression, where the number of visits for each patient is variant. We instantiate one fully connected layer for each auxiliary output task. For instance, the age and mean deviation predictions have their own fully connected layers consisting of a linear function mapping the 1536 features to one output feature. Each output task shares the convolutional layers from the SNN and ResNet18 backbone; however, they each have their own individual task-specific layers that predict a task-specific output.

To compute the loss for the binary glaucoma progression task, we used a binary cross entropy loss (BCE) [14] with a built-in non-linear sigmoid activation function. To address the slightly class imbalanced progressing/non-progressing class proportions, we set the positive class with 3.18 more weight when computing the loss. This value was derived by the ratio between the number of negative samples vs. the number of positive samples in our training dataset. This artificially "augments" the positive class in a way to ensure our model is learning both classes equally. Because the mean deviation, age, global thickness, and OCT clock-hour data are all continuous values, we used a mean squared error (MSE) loss function for all of the auxiliary outputs.

$$BCE = -(y \log(p) + (1 - y) \log(1 - p))$$
$$MSE = \sum_{i=1}^{D} (x_i - y_i)^2$$
$$LOSS = (-(y \log(p) + (1 - y) \log(1 - p))) * \lambda_1 + (\sum_{i=1}^{D} (x_i - y_i)^2) * \lambda_2)$$

To combine the auxiliary loss with the main progression task loss, we simply summed each of the individual losses before computing the backwards propagation. Each loss is weighted with λ, a floating point value between 0 and 1 that weighs the importance of the task. In previous research, it has been shown that structural glaucoma findings might have a greater effect on determining glaucoma progression[15]. Knowing this, we weighted the auxiliary losses of age and mean deviation (0.3) to have a higher loss weight than (0.2) to apply more "attention" to these tasks when computing the loss. The values 0.2 and 0.3 are given so that we can maintain a larger emphasis on the main glaucoma progression loss (0.7 or 0.8). As shown in the equation above, we multiply the main BCE loss with λ_1, which should hold the highest loss weight. The second half of the formula represents the specific auxiliary task we are dealing with, and it is a MSE loss function with λ_2 being the weight of the auxiliary loss. The sum of λ_1 and λ_2 should always equate to 1.

3 Experiments and Results

In this study, we ran a comparative analysis amongst the baseline model of predicting glaucoma progression and the models with age, mean deviation, global

thickness, and OCT clock-hour data to see which auxiliary outputs show an improvement in the main glaucoma progression task. To evaluate each model's performance, we used the area under the receiver operating characteristic curve to gain insight on how well our model is predicting true positives and true negatives. We first experimented with the random forest, support vector machine (SVM), Naïve Bayes, and logistic regression classifiers to predict glaucoma progression as justifications for the need for deep learning. These four approaches are commonly used in medical imaging[16, 17]. In this study, the baseline models are used to justify the need for utilizing deep learning and our auxiliary domain learning. In order to ensure a fair comparison between all of our models, we instantiated a seeding algorithm for reproducible weight initializations and repeatable augmentations. The baseline model trained for only 19 epochs at a

Auxiliary Output(s)	**ROC-AUC**	**Loss Weight (λ)**	**Epoch Converged**
Baseline Siamese Neural Network	84.8±4.9%	N/A	19
Age	87.3±4.3%	0.3	74
Mean Deviation	89.1±3.8%	0.3	16
Global Thickness	90.8±3.3%	0.2	102
Global Thickness + Clock Hours	**91.3±2.6%**	**0.2**	**72**

Baseline Models	**ROC-AUC**	**Loss Weight (λ)**	**Epoch Converged**
Random Forest [18]	86.6±4.6%	N/A	N/A
SVM [19]	81.8±4.6%	N/A	N/A
Naïve Bayes [20]	81.9±3.8%	N/A	N/A
Logistic Regression	84.1±4.3%	N/A	N/A

Table 1: ROC-AUC scores for each auxiliary output. Each loss was weighted with λ shown on the third column. We illustrate the strong regularization that each auxiliary output provides.

learning rate of 1×10^{-4} before the validation ROC-AUC stopped improving. At this stage, the ROC-AUC result for the test set was 84.8%. When we added age as an auxiliary loss with a weight $\lambda = 0.3$, the validation ROC-AUC reached its peak at 74 epochs allowing for longer training, and in result a better ROC-AUC on the test set (87.3%). Although the MD auxiliary output converged even faster than the baseline model, it is evident that the features of mean deviation were learned through the 16 epochs because the testing ROC-AUC increased to 89.1%. The features that improved the ROC-AUC performance the most on the testing set was the OCT data. Just including one continuous output (global thickness) improved the baseline model to 90.8% ROC-AUC, and finally when concatenating the global thickness to the twelve-clock-hour vector as an additional output, the network yielded an ROC-AUC of 91.3%. To obtain an error margin for each of our model, we utilized a bootstrapping sampling method, where a random observation is chosen from our dataset and scored using ROC-AUC.

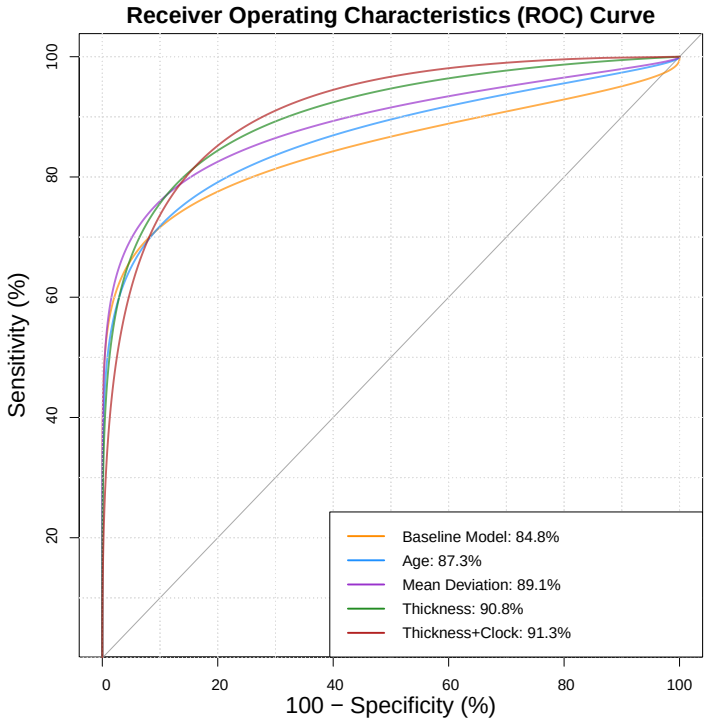

Fig. 3: Illustration of the receiver operating characteristic (ROC) curve for our baseline and multi-output networks. We performed a binormal fitting to smoothen each ROC curve.

4 Discussion

The end goal for this project was to train a convolutional neural network to capture relevant features to detecting glaucoma progression through multi-domain learning, where auxiliary tasks are solved alongside the main task. We noticed that when giving the network ground truth values such as the age, mean deviation, and OCT data, there was a significant increased in ROC-AUC. As shown in the ROC-AUC curves in Figure 3, the ROC-AUC improved from 84.8% to 91.3% after adding relevant output labels. This is crucial because the neural network captured relevant information of the auxiliary labels without needing any input other than the optic disc photographs. To ensure that our deep learning model captured the relevant features to predict glaucoma progression, we generated "heat-maps" based on the region based attribution method (XRAI) [21] to localize the areas of the neural network with the steepest gradients in respect to the second optic disc photograph (Figure 4). In application, the auxiliary-domain neural network can be directly implemented by clinicians by inputting the first three optic disc photographs into the model, and it will predict glaucoma progression by utilizing already learned features from OCT, age, and MD data. We

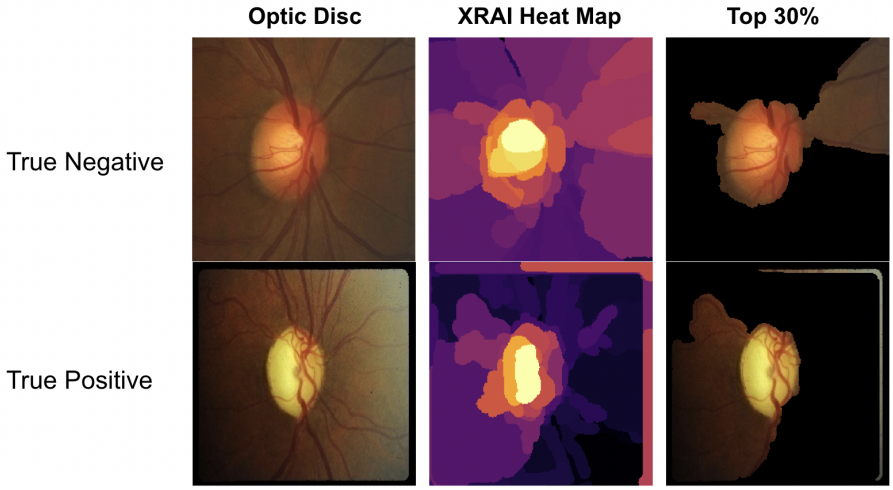

Fig. 4: XRAI heat maps depicting true positive and true negative glaucoma progression predictions. This ensures that our model is learning relevant features to glaucoma progression.

also noticed that auxiliary outputs act as a heavy regularization method that helps the model train for more epochs without overfitting. As shown in Table 1, the baseline no auxiliary output network converged after just 19 epochs. However, a majority of the auxiliary outputs took more than fifty epochs to reach the optimum solution. For example, the progression + global thickness model was trained for 102 epochs before reaching an optimum solution. We did not include all auxiliary outputs combined in one model to ensure that the majority of the loss was still allocated to the main progression prediction task.

There are several methods to further improve our multi-output neural network. One potential method is implementing is an encoder-decoder U-Net[22] model that learns to segment the optic disc and vessels from each optic disc photograph. By doing so, we hope that the model encapsulated important structural information related to glaucoma that is found in the optic disc and vessel. Because manually labeling the masks for vessels and optic discs is a timely task, we plan to pre-train a segmentation model from the open source FIVES dataset[23] and Refuge Challenge[24] dataset, and apply the pre-trained model on our dataset for pseudo ground truth labels.

5 Conclusion

This paper describes a new training pipeline for auxiliary-domain learning for an improved prediction of glaucoma progression. We used hard parameter sharing between the final fully connected layers for each task, and created a combined loss function that weights both the main task, and each auxiliary task with different values of λ scaled between 0 and 1 to balance the importance of the

task during back propagation. One limitation to our study may be a lack of clinical data such as pattern standard deviation, gender, and the visual field count that can even further improve the main glaucoma progression prediction. Our model performs well on non-progressing eyes; however, there are still a few false positives due to our weighted loss function that can be fixed by using other class-balancing methods such as random under or oversampling[25]. We conclude that an auxiliary-domain network leads to a significant increase in ROC-AUC for the main classification task when learning to predict relevant auxiliary features. The auxiliary tasks provide a strong regularization to a convolutional neural network, allowing the model to train for more epochs and generalize better on external test data.

The training pipeline we constructed can handle any form of pairwise or sequential image data for a classification task. We expect that our auxiliary-domain pipeline is a novel contribution that can be used by many to solve machine learning problems from a wide range of research topics across various disciplines, including biology, physics, and medicine. The auxiliary-domain learning that we demonstrated in this study shows promise to have a large impact for medical diagnosis in regions around the globe that may not have access to state of the art machinery such as an OCT scanner to assess the health of a patient. Rather than needing expensive machinery, researchers and clinicians can simply pre-train a multi-output convolutional neural network similar to ours, and deploy the model on readily available data.

References

1. Quigley, H.A., Broman, A.T.: The number of people with glaucoma worldwide in 2010 and 2020 (2006)
2. Allison, K., Patel, D., Alabi, O.: Epidemiology of glaucoma: The past, present, and predictions for the future (2020)
3. Ting, D.S., Peng, L., Varadarajan, A.V., Keane, P.A., Burlina, P.M., Chiang, M.F., Schmetterer, L., Pasquale, L.R., Bressler, N.M., Webster, D.R., et al.: Deep learning in ophthalmology: the technical and clinical considerations. Progress in retinal and eye research **72** (2019) 100759
4. Abràmoff, M.D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J.C., Niemeijer, M.: Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investigative ophthalmology & visual science **57** (2016) 5200–5206
5. Shibata, N., Tanito, M., Mitsuhashi, K., Fujino, Y., Matsuura, M., Murata, H., Asaoka, R.: Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Scientific reports **8** (2018) 14665
6. Christopher, M., Bowd, C., Belghith, A., Goldbaum, M.H., Weinreb, R.N., Fazio, M.A., Girkin, C.A., Liebmann, J.M., Zangwill, L.M.: Deep learning approaches predict glaucomatous visual field damage from oct optic nerve head en face images and retinal nerve fiber layer thickness maps. Ophthalmology **127** (2020) 346–356
7. Dixit, A., Yohannan, J., Boland, M.V.: Assessing glaucoma progression using machine learning trained on longitudinal visual field and clinical data. Ophthalmology **128** (2021) 1016–1026

8. Hemelings, R., Elen, B., Barbosa-Breda, J., Blaschko, M.B., De Boever, P., Stalmans, I.: Deep learning on fundus images detects glaucoma beyond the optic disc. Scientific Reports **11** (2021) 20313
9. Sun, Y., Zhu, L., Wang, G., Zhao, F.: Multi-input convolutional neural network for flower grading. Journal of Electrical and Computer Engineering **2017** (2017)
10. Koch, G., Zemel, R., Salakhutdinov, R., et al.: Siamese neural networks for one-shot image recognition. In: ICML deep learning workshop. Volume 2., Lille (2015)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR **abs/1512.03385** (2015)
12. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. (2009) 248–255
13. Caruana, R.: Multitask learning: A knowledge-based source of inductive bias1. In: Proceedings of the Tenth International Conference on Machine Learning, Citeseer (1993) 41–48
14. Zhang, Z., Sabuncu, M.R.: Generalized cross entropy loss for training deep neural networks with noisy labels. CoRR **abs/1805.07836** (2018)
15. Nouri-Mahdavi, K., Mohammadzadeh, V., Rabiolo, A., Edalati, K., Caprioli, J., Yousefi, S.: Prediction of visual field progression from oct structural measures in moderate to advanced glaucoma. American journal of ophthalmology **226** (2021) 172–181
16. Sarica, A., Cerasa, A., Quattrone, A.: Random forest algorithm for the classification of neuroimaging data in alzheimer's disease: a systematic review. Frontiers in aging neuroscience **9** (2017) 329
17. Lo, C.S., Wang, C.M.: Support vector machine for breast mr image classification. Computers & Mathematics with Applications **64** (2012) 1153–1162
18. Breiman, L.: Random forests. Machine learning **45** (2001) 5–32
19. Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B.: Support vector machines. IEEE Intelligent Systems and their applications **13** (1998) 18–28
20. Rish, I., et al.: An empirical study of the naive bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence. Volume 3. (2001) 41–46
21. Kapishnikov, A., Bolukbasi, T., Viégas, F., Terry, M.: Xrai: Better attributions through regions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2019) 4948–4957
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. CoRR **abs/1505.04597** (2015)
23. Jin, K., Huang, X., Zhou, J., Li, Y., Yan, Y., Sun, Y., Zhang, Q., Wang, Y., Ye, J.: Fives: A fundus image dataset for artificial intelligence based vessel segmentation. Scientific Data **9** (2022) 475
24. Orlando, J.I., Fu, H., Breda, J.B., Van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical image analysis **59** (2020) 101570
25. Mohammed, R., Rawashdeh, J., Abdullah, M.: Machine learning with oversampling and undersampling techniques: overview study and experimental results. In: 2020 11th international conference on information and communication systems (ICICS), IEEE (2020) 243–248