Enhancing ResNet with Ghost Weight Normalization For Improved Retina Disease Classification

Galih Restu Baihaqi[1*], Shafatyra Reditha Shalsadilla[2], Maulida Khairunisa Argaputri[3]

[1,2,3]Faculty of Computer Science, Brawijaya University, Indonesia

Article Information

Received: 21-11-2024
Revised: 28-11-2024
Published: 05-12-2024

Keywords
Classification; Ghost Weight Normalization; ResNet; Retina

****Correspondence Email:***
236150100111027@student.ub.ac.id

Abstract

Retinal disease is a dangerous disease. If left untreated, it can cause blurred vision and even cause permanent blindness. Recently, deep learning approaches are widely used to classify medical diseases. A widely used model to classify medical diseases is ResNet. To train the ResNet model, the data used is data obtained from Kaggle with the name Retinal OCT Images (Optical Coherence Tomography) consisting of 4 classes namely choroidal neovascularization (CNV), DRUSEN, diabetic macular edema (DME), and Normal with a total of 83,600 data. The ResNet base model showed accuracy and f1-score of 92%. Modifying the ResNet Base model with the addition of Ghost Weight Normalization (GWN) which aims to provide more weight normalization opportunities shows an increase in accuracy and f1-score to 94%. GWN can also increase the accuracy of CNN Base from 77% to 81%. This improvement shows that GWN can improve the accuracy of Deep learning models with its weight normalization variation technique. Although the training load and training time when using GWN can increase, the accuracy and f1-score of the ResNet model with GWN of 94% can make the chance of misclassification of retinal diseases smaller.

1. Introduction

The retina is an important part of the human eye that captures incoming light and converts it into signals that are sent to the brain via the optic nerves (Rea et al., 2021). The health of the retina greatly affects a person's vision, and damage to the retina can lead to decreased vision quality and blindness. Some diseases of the retina, such as choroidal neovascularization (CNV), DRUSEN, and diabetic macular edema (DME), are very dangerous conditions (Kim, 2022). The process of diagnosing retinal diseases using conventional medical methods often takes a long time and is costly, which can be an obstacle for patients to get timely treatment (Liu et al., 2020). Therefore, a faster and more efficient approach is needed to detect retinal diseases early to support more effective and affordable treatment.

Recent advances in the field of artificial intelligence, especially deep learning, have opened up many great opportunities in medical image processing, including retinal disease classification. Convolutional Neural Network (CNN) is part of a method in deep learning that is widely used to process many cases (Manjunath &

Mangali, 2023), for example in processing medical image cases. One of the development models of CNN is Residual Network (ResNet). ResNet has the ability to overcome the problem of missing gradients during the training process through a residual learning approach (Yassin et al., 2021). ResNet has been widely used in performing processes on medical images. For example in research, ResNet is used in the classification process of medical images to obtain 71% accuracy (Ying, 2022). Another study on retinal diseases obtained 93% accuracy with another CNN model, namely Google's Inception v3 with transfer learning (Roy Kyameliaand Chaudhuri, 2020.

Although ResNet offers high accuracy in classifying images, there are still challenges in improving its prediction quality, especially in retinal disease classification tasks. One of the main obstacles is the difficulty in handling high data variability and weight complexity. The complexity of the weights can cause the model to struggle to learn optimally, resulting in decreased generalization to new data (Xu & Wang, 2020; Zhang et al., 2020). On the other hand, models are also often limited to one type of weight normalization, which makes them not flexible enough to reflect complex data characteristics.

To overcome this limitation, the Ghost Weight Normalization (GWN) method is used as a technique in weight normalization. GWN is a weight normalization technique that provides many combinations of weight normalization in the network, which allows the model to explore more diverse weight configurations during the training process (Baihaqi & Setiawan, 2024). It works by dividing the whole weight into several parts, then the parts are normalized individually, and then combined back into the original weight. This approach helps the model find more optimal weights. By providing the opportunity for a wider variety of normalizations, GWN can support models in capturing complex patterns in medical data, including retinal images.

Therefore, in this study, Ghost Weight Normalization (GWN) is used to improve the performance of the ResNet model by providing various combinations of weight normalization, allowing the model to be more adaptive in capturing complex patterns in retinal medical images. This research is expected to make a significant contribution to the world of health, especially in supporting the development of retinal medical image processing technology for more accurate and efficient disease detection and classification.

2. Literature Review

2.1 Retina

The retina is a complex structure located at the back of the eyeball and plays an important role in the vision process. The retina is a thin layer of nerve tissue capable of detecting light and converting it into signals that are passed on to the brain via the optic nerve. The retina functions like a camera sensor, capturing light focused into the lens of the eye and converting it into images that can be interpreted by the brain. The main components of the retina are photoreceptors which are rod and cone cells (Azimipour et al., 2020).

The retina is one of the most important indicators to assess the health of the eyes and the body in general. Various retinal disorders can cause visual impairment (Haddad et al., 2023). Common retinal diseases are Choroidal Neovascularization (CNV), a pathological condition that occurs in the retina when abnormal new blood vessels grow from the choroidal layer through Bruch's membrane into the retina, these blood vessels tend to be fragile and break easily, which can cause fluid leakage into the retinal fluid. Then Diabetic Macular Edema (DME) is a serious retinal disease, which is an eye condition that usually occurs in people with diabetes. DME occurs when small blood vessels in the retina, especially around the macula, leak fluid and protein, causing swelling of the macula. Another disease is DRUSEN, which is a small yellow deposit that forms under the retina, especially in the Bruch's membrane layer. DRUSEN is often considered an early sign of aging of the eye (Do et al., 2023). All the retinal diseases mentioned are shown in Fig. 1.

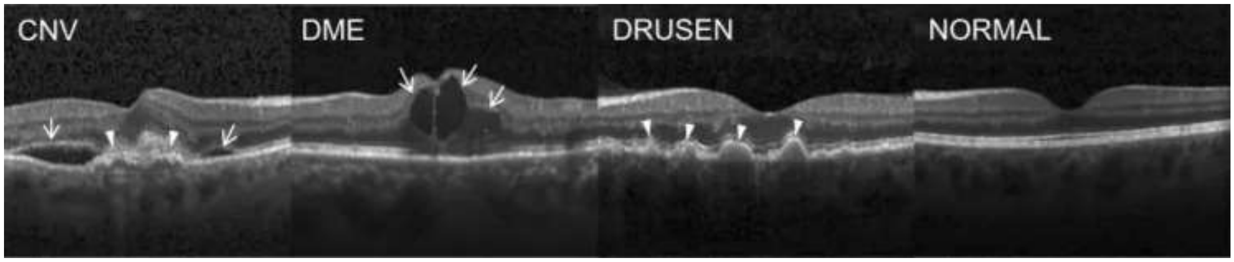

Fig. 1 Retinal Disease

2.2 ResNet

ResNet (Residual Network) is a deep learning architecture designed to overcome vanishing gradient, which is the loss of gradient information when the network becomes very deep. In traditional deep learning, adding layers often leads to performance degradation, causing the accuracy of the model to not improve and may even decrease. ResNet uses the concept of residual learning, allowing the model to focus more on learning the difference between the input and the target output rather than mapping the input to the output directly. Residual block is the main element of ResNet that has shortcut connections that pass the original input directly to the output after going through several layers, making the gradient easier to continue without losing important information (Zahisham et al., 2020).

The Residual Block in ResNet is shown in Fig. Input (x) is the input received by the residual block. F(x) is a non-linear transformation of input x applied through several layers in the residual block and Identity Shortcut is a direct path through the transformation layer connecting input x to the output and the output is the result of the sum of the shortcut connection (x) and the transformation output.

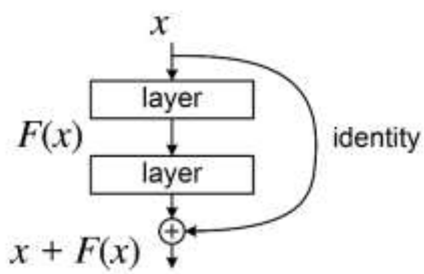

Fig. 2 Residual Block

2.3 Ghost Weight Normalization

Ghost Weight Normalization (GWN) is a normalization technique that offers many variations of normalization on whole weights. The way it works is to divide the degenerate whole weight randomly into several parts according to the size of the ghost size denoted by k so that there will be many combinations of normalized weights in one whole weight. After each weight has been divided into several parts called Ghost Weight, then each Ghost Weight is normalized (Baihaqi & Setiawan, 2024). The normalization formula used is shown in Eq. 1. After the normalization process is complete, each Ghost Weight is merged (Concat) so that it returns to the same whole weight as before. The result of the GWN process is multiplied by the weight of the previous layer which is the input of GWN. Furthermore, this weight will be used by the next layer as in the architecture shown in Fig. 3.

$$Normalization = \sqrt{\left(\sum_{i=0}^{n-1}(Data[i])^2\right)} \quad (1)$$

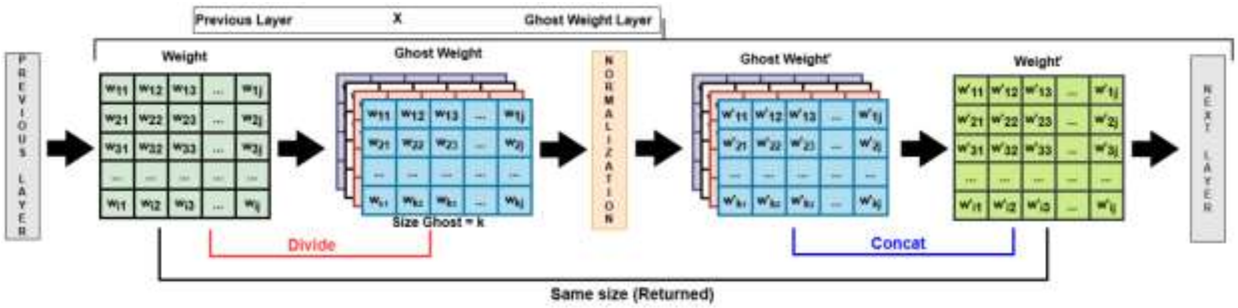

Fig 3. Ghost Weight Normalization Architecture

3. Methods

3.1 Dataset

The dataset used in this study comes from Kaggle under the name Retinal OCT Images (Optical Coherence Tomography), which contains a total of 83,600 retinal image data taken using an OCT device. This dataset is classified into four classes, namely CNV (Choroidal Neovascularization), DME (Diabetic Macular Edema), Drusen, and Normal, each of which represents a specific condition of the retina. The distribution of data in the dataset shows imbalance, where the CNV class is the majority class with 37,216 data or 44.5% of the total dataset. Meanwhile, the Normal class has 26,344 data (31.5%), followed by the DME class with 11,420 data (13.7%) and the Drusen class with 8,620 data (10.3%). This distribution shows that the dataset has an uneven distribution among its classes. Each class has unique visual characteristics, which can help in the model training process to distinguish between normal and abnormal retinal conditions. A visualization of the data distribution between classes is shown in Fig. 4 (b), while an example image from each class is presented in Fig. 4 (a), giving an idea of the variation in this dataset.

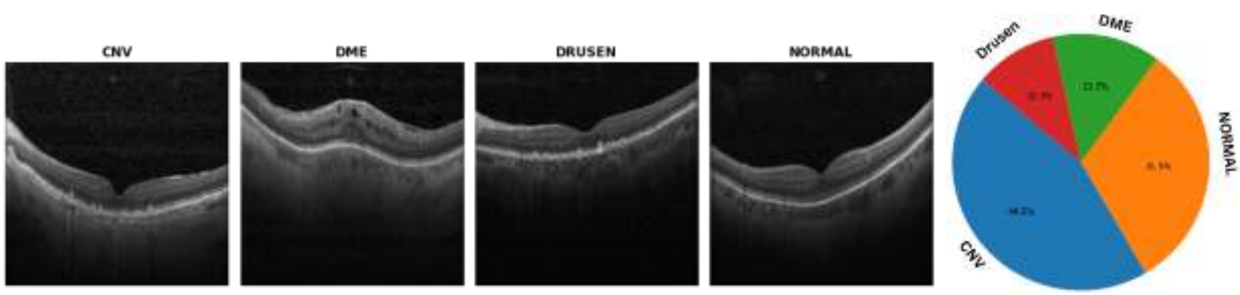

Fig. 4 (a) Dataset, (b) Percentage of Total Data

3.2 Model Architecture

GWN on ResNet architecture is applied to replace ResNet's main architecture, Batch Normalization (BN). BN is one of the techniques used in neural networks that aims to improve the stability and speed of training by normalizing the input for each layer. In this research, BN is replaced using GWN to further introduce variations on the normalization technique in each weight generated in the first convolution layer for the next convolution process. The architecture of the proposed model is a modified ResNet with a GWN change in the BN layer shown in Fig. 5.

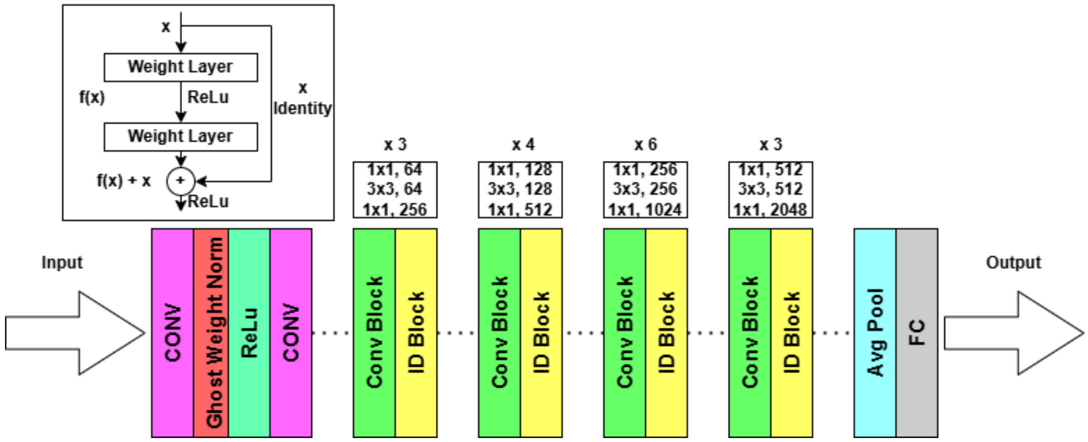

Fig. 5 Model Architecture

3.3 Model Evaluation

Confusion Matrix is a technique used to evaluate the performance of classification models. This matrix shows the comparison between the predictions generated by the model and the ground truth values of the dataset. The Confussion Matrix aims to understand the extent to which the classification model is able to distinguish each class, especially for datasets with more than one class and the display of the Confussion Matrix is shown in Fig. 6. In the medical world, accuracy is used to evaluate the model's ability to predict correctly based on the overall data and F1-Score is also important because it combines Precision and Recall to measure the model's performance. To calculate accuracy is shown in Eq. 2 and for F1-Score is shown by Eq. 3, where TP is True Positive, FP is False Positive, TN is True Negative, and FN is False Negative.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$F1 = 2 \times \frac{\frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \tag{3}$$

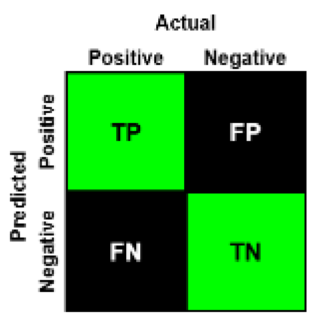

Fig. 6 Confussion Matrix

4. Result and Discussion

In general, the flow of system work starts from the dataset obtained from Kaggle which is then preprocessed. Data preprocessing is done by down sampling the dataset to equalize the majority class so that the total data is the same as the minority class. In addition, the image is also resized to 224x224. Then, the training process is carried out using the ResNet model that has been modified by changing BN to GWN. After the model has finished training, the model is evaluated using the Confusion matrix to calculate accuracy and f1-score. The flow of the system is shown in Fig. 7.

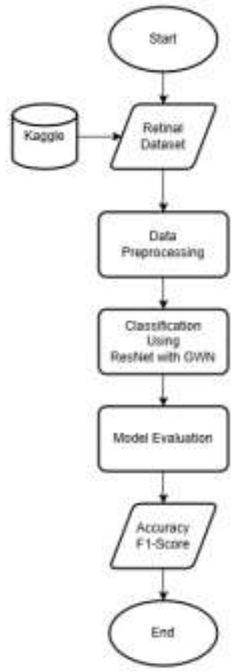

Fig. 7 System Workflow

4.1 Down Sampling Dataset

Fig. 4 (b) shows that the dataset is not balanced so it is necessary to do under sampling so that during the training process, the model does not overfit the majority class. In the dataset, the minority class is DRUSEN with 8,620 data so that other classes such as CNV, Normal, and DME are only used as much as 8,620 data, so that the data processed by the model is equal to the total of all processed data is 34,480 and shown in Fig. 8. Furthermore, the data is divided into 3 parts, namely train, val, and test with a data ratio of 80:10:10 so that the train data is 27,584 data. Then for each val and test data as much as 3,448 data.

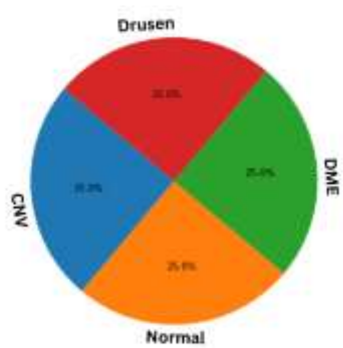

Fig. 8 Distribution of Datasets by Class After Down Sampling

4.2 Proposed Method Result

During the training process, the epoch used to train is 30 epochs with a batch size of 64. Accuracy on validation data during the training process shows that the highest value is at the 26th epoch which indicates that this value is the best epoch during the training process. It can be seen in the graph shown in Fig. 9 (a) that there is a decrease in performance at epoch 30 (last). However, to evaluate the model, the best epoch is 26. When predicting the testing data using a model with the weight of the 26th epoch produces an accuracy and f1-score of 94%. The total training time for 30 epochs is 1,738.14 seconds and the training load is 8,348.16 MB. It can be seen in Fig. 9 (b) that the model can correctly predict 822 CNV class datasets. Then for the DME class, the model managed to predict correctly as many as 815, then followed by the DRUSEN class as many as 800. Meanwhile, the class that was least successfully predicted correctly was the Normal class with a total of 788 data that was successfully predicted correctly.

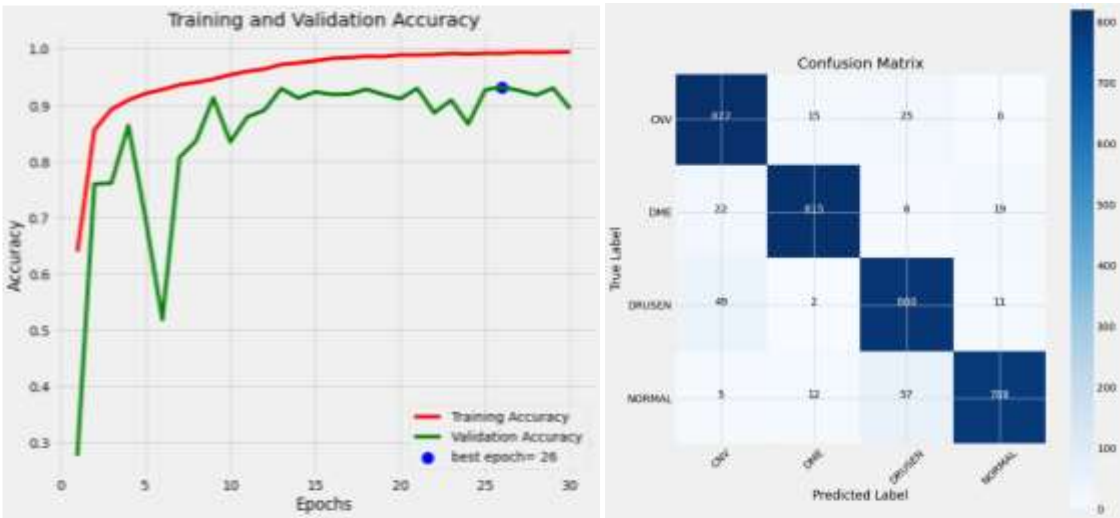

Fig. 9 (a) Training Evaluation Results, (b) Confusion Matrix Result

4.3 Model Performance Comparison

The model was trained using an NVIDIA A100 GPU and overall, the GWN-improved model can improve accuracy and f1-score. For example, in the CNN base, the model arrangement used is the convolution and max pooling layers arranged twice, then the classification layer, resulting in an accuracy and f1-score of 77%. Inserting one GWN layer between each convolution and max pooling layer showed a 4% increase in accuracy and f1-score. In the ResNet base model as well, the increase in accuracy and f1-score when inserting GWN to replace BN can improve the results by 2%. Although the overall results can be seen in Table 1, when using GWN, it can increase the training load and training time. The highest training load and training time is obtained by the CNN model with GWN because the model architecture applies 2 GWNs, but it can also be confirmed that without GWN, the entire model can be lighter and faster. Overall, the best model is ResNet with GWN with 94% accuracy and precision with a training load of 8,348.16 MB and a training time of 1,738.14 seconds as shown in Table 1.

Table 1. Model Perfomance Comparison

No	Model Name	Accuracy (%)	F1-Score (%)	Training Load (MB)	Training Time (Sec)
1	CNN Base	77	77	7,875.45	1,768.91
2	CNN + GWN	81	81	8,577.84	3,939.56
3	ResNet Base	92	92	8,288.42	1,611.17
4	**ResNet + GWN (Proposed Method)**	**94**	**94**	**8,348.16**	**1,738.14**

5. Conclusions

Retinal disease is a dangerous disease because it can cause visual impairment and can lead to blindness. Recently, deep learning approaches are widely used for early diagnosis of diseases, especially retinal diseases. GWN shows excellent performance in improving the accuracy and f1-score of deep learning models. On CNN, GWN improves accuracy and f1-score by 4% compared to CNN base. While on ResNet, GWN improves accuracy and f1-score by 2% compared to ResNet base. Although the use of GWN in CNN and ResNet can increase the training load and training time, the higher accuracy and f1-score indicate that the error rate in predicting retinal diseases is getting smaller. Overall, the ResNet and GWN models performed very well with 94% accuracy and f1-score. In the future, GWN can be implemented on other CNN models such as DenseNet, InceptionNet, and other CNN models.

6. References

Azimipour, M., Valente, D., Vienola, K. V, Werner, J. S., Zawadzki, R. J., & Jonnal, R. S. (2020). Optoretinogram: optical measurement of human cone and rod photoreceptor responses to light. *Optics Letters*. https://doi.org/10.1364/ol.398868

Baihaqi, G. R., & Setiawan, B. D. (2024, October). Ghost Weight Normalization on Wind Speed Prediction Model Using Gated Recurrent Unit. *2024 International Conference on Sustainable Information Engineering and Technology (SIET) (SIET 2024)*.

Do, M. T., Huynh, H. N., Tran, T. N., & Hoang, T. (2023). Prediction of Retina Damage in Optical Coherence Tomography Image Using Xception Architecture Model. *2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, 58–61. https://doi.org/10.1109/ECBIOS57802.2023.10218586

Haddad, Z., Yaya, B. M., Zgolli, H., Sidibé, D., Tabia, H., & Khlifa, N. (2023). Retinal pathologies detection in OCT images based on Bilinear convolutional neural network. *2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA)*, 1–7. https://doi.org/10.1109/INISTA59065.2023.10310341

Kim, J. E. (2022). Exploring neovascular age-related macular degeneration and diabetic macular edema and advances in treatment. *The American Journal of Managed Care*, *28 3 Suppl*, S35–S43. https://doi.org/10.37765/ajmc.2022.88853

Liu, J. C., Gibson, E., Ramchal, S., Shankar, V. A., Piggott, K., Sychev, Y., Li, A. S., Rao, P., Margolis, T., Fondahn, E., Bhaskaranand, M., Solanki, K., & Rajagopal, R. (2020). Diabetic Retinopathy Screening with Automated Retinal Image Analysis in a Primary Care Setting Improves Adherence to Ophthalmic Care. *Ophthalmology. Retina*. https://doi.org/10.1016/j.oret.2020.06.016

Manjunath, M., & Mangali, K. (2023). Development and deployment of advanced machine learning frameworks for the identification,analysis,and neutrolzation of botnet activities throught extensive network traffic. *The 4th International Conference on Information Technology and Security*, 167–173.

Rea, M., Nagare, R., & Figueiro, M. (2021). Modeling Circadian Phototransduction: Retinal Neurophysiology and Neuroanatomy. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.615305

Roy Kyamelia and Chaudhuri, S. S. and R. P. and C. S. and B. S. (2020). Transfer Learning Coupled Convolution Neural Networks in Detecting Retinal Diseases Using OCT Images. In S. Mandal J. K. and Banerjee (Ed.), *Intelligent Computing: Image Processing Based Applications* (pp. 153–173). Springer Singapore. https://doi.org/10.1007/978-981-15-4288-6_10

Xu, Y., & Wang, X. (2020). Weight normalized deep neural networks. *Stat*, *10*. https://doi.org/10.1002/sta4.344

Yassin, W., Abdollah, M. F., Muslim, Z., Ahmad, R., & Ismail, A. (2021). An Emotion and Gender Detection Using Hybridized Convolutional 2D and Batch Norm Residual Network Learning. *Proceedings of the 2021 9th International Conference on Information Technology: IoT and Smart City*. https://doi.org/10.1145/3512576.3512590

Ying, Z. (2022). Classification of Diabetic Retinopathy Based on B-ResNet. *2022 7th IEEE International Conference on Data Science in Cyberspace (DSC)*, 570–575. https://doi.org/10.1109/dsc55868.2022.00085

Zahisham, Z., Lee, C., & Lim, K. (2020). Food Recognition with ResNet-50. *2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 1–5. https://doi.org/10.1109/IICAIET49801.2020.9257825

Zhang, G., Niwa, K., & Kleijn, W. (2020). Projected Weight Regularization to Improve Neural Network Generalization. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4242–4246. https://doi.org/10.1109/ICASSP40776.2020.9054133