RESEARCH ARTICLE

# Performance Analysis of Efficient Deep Learning Models for Multi-Label Classification of Fundus Image

**Muhammed Pektaş** [a†]

[a] İzmir Bakırçay University, İzmir, Turkey
† mhmdpkts@gmail.com, corresponding author

**RECEIVED** AUGUST 8, 2023
**ACCEPTED** SEPTEMBER 7, 2023

**CITATION** Pektaş, M. (2023). Performance analysis of efficient deep learning models for multi-label classification of fundus image. *Artificial Intelligence Theory and Applications, 3(*2), 105-112.

---

**Abstract**

Convolutional Neural Networks (CNNs) have demonstrated significant advancements in the domain of fundus images owing to their exceptional capability to learn meaningful features. By appropriately processing and analyzing fundus images, computer-aided diagnosis systems can furnish healthcare practitioners with valuable reference information for clinical diagnosis or screening purposes. Nevertheless, prior investigations have predominantly concentrated on detecting individual fundus diseases, while the simultaneous diagnosis of multiple fundus diseases continues to pose substantial challenges. Furthermore, the majority of previous studies have prioritized diagnostic accuracy as their main focus. Efficient Deep Learning constitutes a crucial concept that enables the utilization of deep learning models on edge devices, thereby reducing the computational carbon footprint. Facilitating the cost-effective diagnosis of eye diseases from fundus images on edge devices holds significance for researchers aiming to deploy these vital healthcare models into practical use. This study focuses on assessing the performance of well-known efficient deep learning models in addressing the multi-label classification problem of fundus images. The models underwent training and testing using the dataset provided by international competition on ocular disease intelligent recognition in 2019. The experimental findings demonstrate that the efficientnetb3 model outperforms the other models, exhibiting the highest level of performance. And also, when applying standard data augmentation techniques to the current dataset, we observe decreasing in f1-score and accuracy. Code is available at https://github.com/m-pektas/Performance-Analysis-of-Efficient-Deep-Learning-Models-for-MultiLabel-Classification-of-Fundus-Image

**Keywords:** artificial intelligence; deep learning; efficient deep learning; multi-label classification; eye disease diagnosis; fundus images

---

## 1. Introduction

Diseases of the fundus can lead to loss of vision, which is the leading cause of blindness [1]. There are some common diseases that can be diagnosed from fundus images such as cataracts, diabetes, etc. For example, cataract disease is one of the important diseases. An estimated 95 million people worldwide are affected by cataract, according to the World Health Organization (WHO) [2]. In middle and low-income countries, cataract remains the leading cause of blindness. The development of late-stage fundus diseases often has a serious impact on the visual function of patients, and there is no specific treatment for such diseases. Diabetes patients are now one of the largest



Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2023 İzmir Bakırçay University

disease groups in the world. There are no obvious abnormal symptoms in the early stages of the disease. However, it eventually leads to blindness. It is one of the four most common blinding diseases. [3]. If the disease is caught in the early stages, it is still treatable. If diagnosed at a late stage, the visual prognosis is very poor and the cost of treating is high, even if surgery is successful [4]. The patient's vision will rapidly deteriorate within a short period of time, resulting in irreversible visual impairment, if prompt and effective treatment is not provided [5].

Early detection and treatment of fundus diseases are important. Artificial Intelligent technology has the potential to assist primary care ophthalmologists in diagnosing ocular diseases by leveraging comprehensive medical data, thereby offering new strategies to improve the ability to diagnose and treat in primary care clinics. Integrating artificial intelligence into ophthalmic practice promises to meet the practical needs of many patients suffering from fundus-related diseases. Therefore, developing and deploying efficient and accurate machine learning models plays a crucial role in enabling early diagnosis, improving the availability of healthcare services, and making them more cost-effective.

In previous works, researchers designed and implemented deep learning models to diagnose eye diseases in the early stage of the treatments. In this paper, we analyze the efficient CNN models' performance on the ocular disease intelligent recognition (ODIR-5K) dataset due to care about cost and efficiency in health applications to increase accessibility of health technologies.

In this paper, we analyze previous works about eye disease diagnosis from fundus image in Section 2. In Section 3, we give information about dataset and methods that we used. In Section 4, we share our experiment results.

## 2. Literature Review

Tham et al. work on diagnosing cataract disease. They compare AI algorithm performance and clinician performance on this problem. Using more than 25,000 images from population-based studies, they report the development and validation of a retinal photograph-based deep learning algorithm for automated detection of visually significant cataract [6]. Wang et al, work on the ODIR-5K dataset and focus on diagnosing multi-disease classification from fundus image problems about the eye. They propose a method that uses the EfficientNets [7] model with ensemble learning [8]. In Zhou et al’s work, they focus on diabetic retinopathy. As they claim that, most researchers can’t obtain satisfactory performance on diabetic retinopathy due to less training data that is without consistent annotation. Therefore, they demonstrate a dataset that has consisted of pixel and image level annotations [9]. He et al, propose a new method that uses attention and different fusion techniques. They achieve the best performance using Resnet-101 with their proposed fusion technique [10]. Bhati et al, propose a method called Discriminative Kernel Convolution Network (DKCNet) that explores discriminative region-wise features without adding extra computational cost [11]. Wang et al, combine convolutional neural networks and self-attention called MBSaNet for fundus disease identification. Their experimental results show that MBSaNet achieves state-of-the-art performance with fewer parameters than current methods [12].

# 3. Material and Methods

## 3.1. Dataset

The ODIR-5K dataset [13] was obtained from the "International Competition on Ocular Disease Intelligent Recognition" sponsored by the Peking University. Dataset consists of "real" patient data from Shanggong Medical Technology Co. Ltd. from different hospitals/medical centers in China. The training set is a structured ophthalmology database containing 3,500 patients, their ages, left and right eye color fundus images, and doctors' diagnosis codes. Color fundus photographs of 500 patients without age and sex were used as the test set. The fundus images are taken with different image resolutions by various cameras on the market, such as Canon, Zeiss, and Kowa. Thanks to this dataset, the patients can be classified into eight labels. Class names (normal, diabetic retinopathy, glaucoma, cataract, AMD, hypertensive retinopathy, myopia, and other diseases) and distributions are shown in Figure 1.

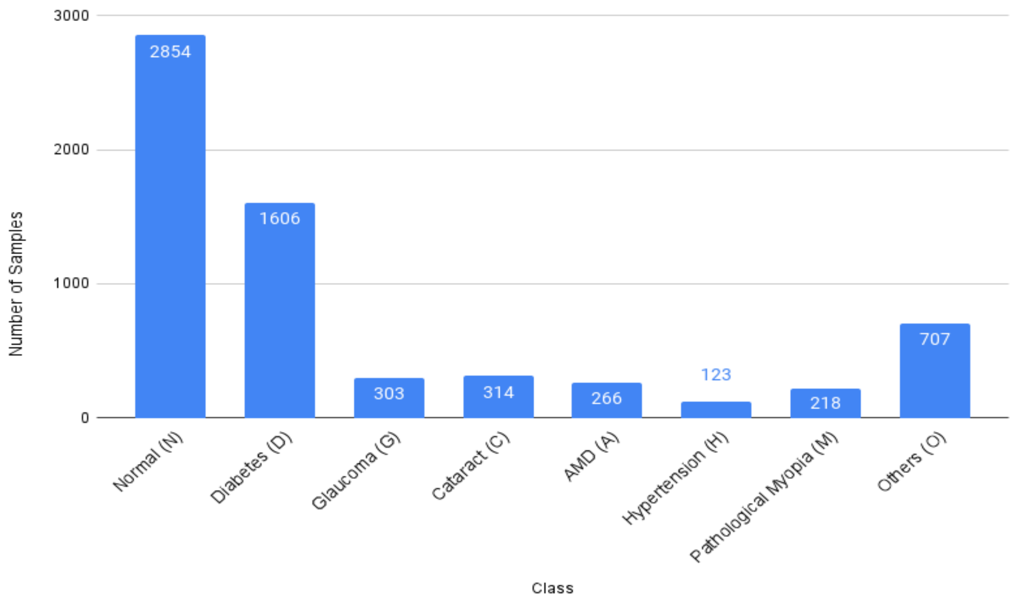


Figure 1. Class names and distribution of ODIR-5K dataset.

## 3.2. Data Pre-processing

In dataset pre-processing step, we used the already pre-processed version of the dataset [14]. The images in this version of the dataset are already resized to 512x512 and cropped by providing 1:1 aspect ratio. The dataset divided into a training set and a test set in a 7:3 ratio, and then resize the images to 224 x 224. This version of the dataset is our baseline.

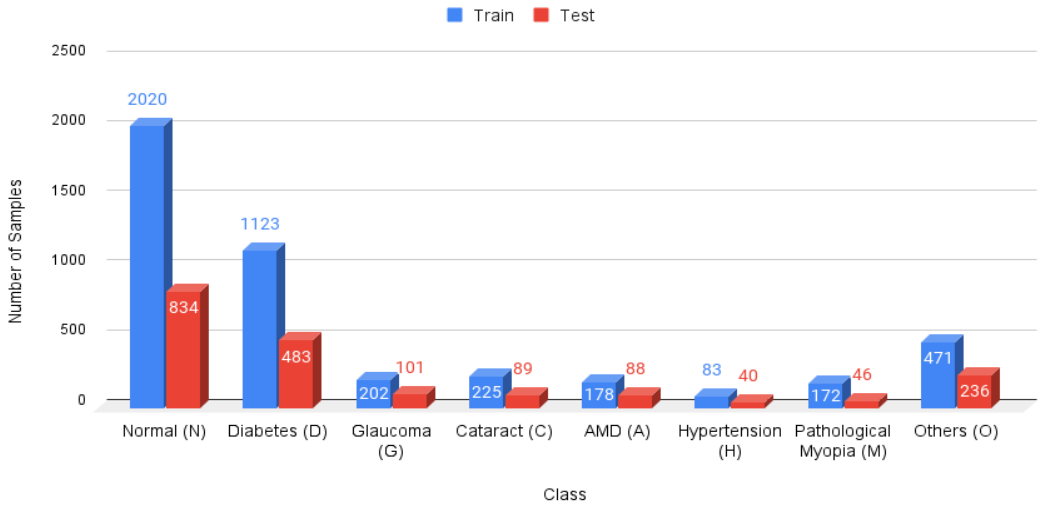


Figure 2. Train and test set distribution of our baseline version of the dataset.

### 3.3. Methods

Mobilenet, EfficientNet, and SqueezeNet versions are well-known architectures in the efficient deep learning community. We use these models and their versions in our experiments. We described these models in the following paragraphs.

***Mobilenets***: MobileNets are developed for resource-constrained use cases. Therefore, it is designed with the focus on making it a small, low-latency, low-power solution. Similar to other popular large-scale models, such as Inception, they can be used for classification, detection, embedding, and segmentation [15].

***EfficientNets***: The baseline network is highly dependent on the effectiveness of model scaling. To further improve performance, they also designed a new baseline network through neural architecture search using the AutoML MNAS framework, which optimizes both accuracy and efficiency (FLOPS). The Efficientnets uses mobile inverted bottleneck convolution as a main module, similar to MobileNetV2 and MnasNet, but is slightly larger due to an increased FLOP budget. As a result, the basic network is then scaled up to obtain a family of networks known as EfficientNets [7, 16].

***SqueezeNets***: SqueezeNet is an algorithm designed for small size and high accuracy. This compact and efficient CNN model was proposed in 2016 by researchers from DeepScale, UC Berkeley, and Stanford University. SqueezeNet balances high accuracy with low complexity, making it ideal for resource-constrained edge devices such as mobile phones. SqueezeNet uses a special type of convolution layer called the fire module, a combination of 1x1 and 3x3 filters, to reduce parameters while maintaining high accuracy, making it efficient for resource-limited devices. [17].

We train these models and their versions on our baseline dataset and we select the best architecture by accuracy and f1-score metrics. After that, we do experiments with different data augmentation and different hyperparameters to obtain the best results. At the end of these experiments, we compare the selected models' accuracy and f1-score performances. We demonstrate that the best performer option is EfficientnetB3 without data augmentation.

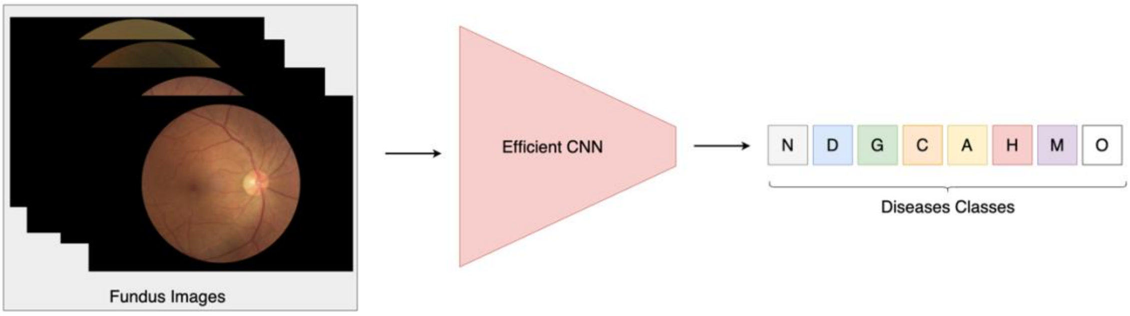


Figure 3. A simple pipeline for multi-label classification of eye diseases with an efficient CNN architecture.

## 4. Experiments and Results

### 4.1. Configurations

During these experiments, we use a computer that has Intel® Core™ i7-7700HQ CPU @ 2.80GHz × 8, NVIDIA GeForce GTX 1050 Ti, and 8 GB ram.

### 4.2. Implementation Details

In these trainings, we use Adam optimizer with 0.001 learning rate. Also, we use ReduceLROnPlateau learning rate schedular to adjust the learning rate during the training process. The batch size and early stopping patient set 15 and 20 respectively. All hyperparameters that we used are shown in Table 1.

Table 1. Hyperparameter configurations.

| Configuration | Value |
|---|---|
| Optimization Function | Adam |
| Maximum Epoch | 100 |
| Batch Size | 15 |
| Learning Rate | 0.001 |
| Learning Rate Schedular | ReduceLROnPlateau |
| Early Stopping Patience | 20 |

### 4.3. Selection of Best Architecture

Firstly, we use our baseline dataset that split with 7:3 train/test ratio during our selection of best efficient architecture experiments. Training set has 4474, testing set 1917 samples with this splitting ratio. In these experiments, we train 10 different efficient deep learning models with the same dataset and hyperparameters. We achieve worst accuracy performances with SqueezeNet versions. Our EfficientNet and MobileNet experiments have competitive accuracy with each other, but best performer model is EfficientNetB3 with 90.12% accuracy.

Based on these results, we select EfficientNetB3 to our further experiments. Experiment results are shown in Table 2.

Table 2. Selection of best architecture experiment results on test dataset.

| Method | Train/Test Split | Best Test Loss | Accuracy |
|---|---|---|---|
| EfficientNetB0 | 0.7 | 0.4629 | 88.12 |
| EfficientNetB1 | 0.7 | 0.4825 | 87.96 |
| EfficientNetB2 | 0.7 | 0.4604 | 87.89 |
| ***EfficientNetB3*** | ***0.7*** | ***0.4297*** | ***90.14*** |
| EfficientNet_V2_S | 0.7 | 0.4754 | 88.26 |
| MobilenetV2 | 0.7 | 0.5835 | 85.03 |
| MobilenetV3_S | 0.7 | 0.5798 | 83.55 |
| MobilenetV3_L | 0.7 | 0.4627 | 87.34 |
| Squeezenet_10 | 0.7 | 1.258 | 51.00 |
| Squeezenet_11 | 0.7 | 1.2854 | 49.59 |

As shown in Table 2, We obtain the best results with EfficientNetB3 and the worst result with Squeezenet_11. Most of these models have higher than 85% accuracy on test dataset. The top 3 results belong to EfficientNetB3, EfficientNet_V2_S, and EfficientNetB0. As our experiments, the EfficientNet model's versions have better performance than other methods such as MobilenetV2, MobilenetV3_L, and Squezenet_10.

## 4.4. Optimizing the First Results

We select the best performer model as EfficientNetB3 on eye disease diagnosis problem based on the Table 2. We try different train/test split ratio to achieve best performance. Our experiment results that we try different dataset split ratio are shown in Table 3.

| Method | Train/Test Split | Best Test Loss | Accuracy | F1-Score |
|---|---|---|---|---|
| *EfficientNetB3* | *0.7* | *0.4297* | *90.14* | 0.6619 |
| *EfficientNetB3* | 0.8 | 0.2656 | 0.9387 | 0.6596 |
| ***EfficientNetB3*** | **0.9** | **0.1606** | **0.9664** | **0.6870** |

Table 3: Train/Test split ratio experiment results.

As shown in Table 3, when we used 0.9 as the dataset split ratio, we obtain the best accuracy and f1-score. These results give us some insights about the model that needs more data to learn diagnosing eye diseases.

Table 4. Data augmentation experiments' results.

| Method | Train/Test Split | Data Augmentation | Best Test Loss | Accuracy | F1-Score |
|---|---|---|---|---|---|
| *EfficientNetB3* | **0.9** | No | **0.1606** | **0.9664** | **0.6870** |
| *EfficientNetB3* | 0.9 | Imagenet | 0.3753 | 0.8857 | 0.5430 |
| *EfficientNetB3* | 0.9 | Cifar10 | 0.3288 | 0.8868 | 0.6056 |
| *EfficientNetB3* | 0.9 | Simple | 0.2092 | 0.9441 | 0.6338 |

We select 0.9 as the best train/test split ratio based on Table 3. We used this ratio in our data augmentation experiments. In our experiments, data augmentation have 4 different values as No, Imagenet, Cifar10, and Simple. Imagenet and Cifar10 are the default data augmentation processes for Cifar10 and Imagenet datasets [18]. Simple is defined by us. It contains only different color jittering with brightness 0.5 and hue 0.3 values. As shown in Table 4, applying data augmentation caused the decreasing model performance. Also, by increasing the data augmentation effect, the model performances are getting decrease. According to Table 3 and 4, we can say that, the model need to more data and training.

Table 5. Comparison of CNN performances between Wang et al [8] and our methods.

| Method | Accuracy | Parameter Count (M) |
|---|---|---|
| EfficientNetB3 (Ours) | **0.96** | **12** |
| VGG 16 | 0.86 | 134.7 |
| VGG 19 | 0.86 | 143.7 |
| InceptionV3 | 0.87 | 6.23 |
| EfficientB3 | 0.90 | 12 |
| Resnet50 | 0.86 | 23.9 |

As shown in Table 5, we compare ours and Wang et al [8] results. We achieve better accuracy with 96% on EfficientNetB3 then others. In addition, EfficientNet has lower parameter count then most of other methods while has higher accuracy. These results support to our less costly and more accessible eye disease diagnosis health applications goal.

## 5. Conclusion

Late-stage fundus diseases often cause severe visual impairment and there is no specific treatment to treat such diseases. In the early stages of most diseases, there are no obvious abnormalities, but eventually, it will cause blindness. It is still treatable if caught in the early stages. When diagnosed in late stages, even with successful surgery, the visual prognosis is very poor and the cost of treatment is high. Therefore, it is very important to detect and treat fundus diseases at an early stage. Combining artificial intelligence and ophthalmic medical treatment is expected to meet the practical needs of many patients suffering from fundus diseases. Thus, efficient and accurate machine learning models have an important role in terms of early diagnosis, increasing the accessibility of health services and making them cheaper. In this work, we analyze the efficient deep CNN model's performance on ODIR-5K dataset for the identification of eye diseases from fundus images. According to our experiment results, EfficientNetB3 best performer among the well-known current efficient CNN models. Also, we make empirical studies on EfficientNetB3 using different hyperparameters and data augmentation techniques. In our experiments, our best results 96.94% accuracy and 0.6870 f1-Score. According to us, these findings can make contribution to accurate, efficient, and more accessible health applications. In future works, we use an automatic hyperparameter tuning process to obtain possible better results and generate synthetic data to make the dataset balanced and diverse.

## References


[1] J. B. Jonas, R. R. A. Bourne, R. A. White, S. R. Flaxman, J. Keeffe, J. Leasher, K. Naidoo, K. Pesudovs, H. Price, T. Y. Wong, S. Resnikoff, and H. R. Taylor, ''Visual impairment and blindness due to macular diseases globally: A systematic review and meta-analysis,'' Amer. J. Ophthalmol., vol. 158, no. 4, pp. 808–815, Oct. 2014.

[2] Liu, Y. C., Wilkins, M., Kim, T., Malyugin, B., & Mehta, J. S. (2017). Cataracts. The Lancet, 390(10094), 600-612.

[3] J. L. Leasher, R. R. A. Bourne, S. R. Flaxman, J. B. Jonas, J. Keeffe, K. Naidoo, K. Pesudovs, H. Price, R. A. White, T. Y. Wong, S. Resnikoff, and H. R. Taylor, ''Global estimates on the number of people blind or visually impaired by diabetic retinopathy: A meta-analysis from 1990 to 2010,'' Diabetes Care, vol. 39, pp. 1643–1649, Sep. 2016, doi: 10.2337/dc15- 2171

[4] O. B. Walton, R. B. Garoon, C. Y. Weng, J. Gross, A. K. Young, K. A. Camero, H. Jin, P. E. Carvounis, R. E. Coffee, and Y. I. Chu, ''Evaluation of automated teleretinal screening program for diabetic retinopathy,'' JAMA Ophthalmol., vol. 134, no. 2, pp. 204–209, Feb. 2016, doi: 10.1001/jamaophthalmol.2015.5083.

[5] H. Ye, Q. Zhang, X. Liu, X. Cai, W. Yu, S. Yu, T. Wang, W. Lu, X. Li, H. Jin, Y. Hu, X. Kang, and P. Zhao, ''Prevalence of age-related macular degeneration in an elderly urban Chinese population in China: The Jiangning eye study,'' Investigative Ophthalmol. Vis. Sci., vol. 55, no. 10, pp. 6374–6380, Sep. 2014, doi: 10.1167/iovs.14-14899.

[6] Tham, Y. C., Goh, J. H. L., Anees, A., Lei, X., Rim, T. H., Chee, M. L., ... & Cheng, C. Y. (2022). Detecting visually significant cataract using retinal photograph-based deep learning. Nature Aging, 2(3), 264-271.

[7] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

[8] Wang, J., Yang, L., Huo, Z., He, W., & Luo, J. (2020). Multi-label classification of fundus images with efficientnet. IEEE Access, 8, 212499-212508.

[9] Zhou, Y., Wang, B., Huang, L., Cui, S., & Shao, L. (2020). A benchmark for studying diabetic retinopathy: segmentation, grading, and transferability. IEEE Transactions on Medical Imaging, 40(3), 818-828.

[10] He, J., Li, C., Ye, J., Wang, S., Qiao, Y., & Gu, L. (2020, April). Classification of ocular diseases employing attention-based unilateral and bilateral feature weighting and fusion. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) (pp. 1258-1261). IEEE.

[11] Bhati, A., Gour, N., Khanna, P., & Ojha, A. (2023). Discriminative kernel convolution network for multi-label ophthalmic disease detection on imbalanced fundus image dataset. Computers in Biology and Medicine, 106519.

[12] Wang, K., Xu, C., Li, G., Zhang, Y., Zheng, Y., & Sun, C. (2023). Combining convolutional neural networks and self-attention for fundus diseases identification. Scientific Reports, 13(1), 76.

[13] Peking University International Competition on Ocular Disease Intelligent Recognition (ODIR-2019), [online] Available: https://odir2019.grand-challenge.org/dataset/.

[14] Kaggle. "Ocular Disease Recognition", Access Date: 11.06.2023. https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k

[15] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[16] Google Research. "EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling", Access Date: 11.06.2023. https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html

[17] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360.

[18] Pytorch. "AutoAugment", Access Date: 11.06.2023. https://pytorch.org/vision/master/generated/torchvision.transforms.AutoAugment.html#torchvision.transforms.AutoAugment