Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/46/20241177

Hyperparameter search and performance assessment of convolutional neural networks for image classification

Jianwen Lyu

School of Computer Science, University of Nottingham, Nottingham, NG7 2RD, United Kingdom

scyjl14@nottingham.ac.uk

Abstract. As the fields of deep learning, computer vision gain increasing popularity. Image classification is a fundamental task in computer vision that aims to understand images holistically and classify them into specific categories. Various types of image classification algorithms based on convolutional neural networks (CNNs) have been developed and evaluated on multiple image datasets. During the training of many image classification algorithms, several factors can significantly impact the results and even the learning and training process. These factors include the complexity of the image inputs, the optimizer method used, and the parameter tuning techniques applied. In this article, the author conducted experiments using ResNet-18, a residual network learns residual functions to avoid overfitting, with different combinations of learning rates and optimizers. After analyzing the experimental results, it could be observed that the extent of input complexity can indeed affect the accuracy of the model's results, as well as its convergence behavior during the training process.

Keywords: Residual Network, Optimizers, Convolutional Neural Network.

1. Introduction

Since the rise of deep learning and relevant algorithms, especially the appearance of convolutional neural networks (CNN), image classification with the help of deep learning algorithm has acquired huge success and has become an increasingly popular and prevail topic nowadays. Since there are abundant resources of images used as datasets for classification tasks in the context of real world, different kinds of image classification problems are gradually coming into the sight of researchers for training and evaluating the performance of image classification algorithms. Also, the CNN's availability of drawing out features from complex and high-dimensional visual information apparently accelerated the development of fields such as object localization, object detection and image captioning [1]. A variety of CNN structures, such as AlexNet by Alex Krizhevesky in 2012, VGG models with more complex layer structure, and residual neural network (ResNet) with deeper layers yet less complexity [2-4], are therefore gradually emerged for solving image classification tasks.

During the process of developing deep learning algorithms for image classification tasks, a large number of factors affecting the performance of CNNs were discovered. First, for instance, different image classification tasks clearly have different degrees of complexity for extracting the features inside the image datasets attaching to the tasks. For example, tasks with smaller datasets such as MNIST digit-recognition task can be easier for CNN to recognize, but tasks with larger datasets such as ImageNet

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/46/20241177

may contain much more variability and require larger learning capacity [2]. Second, different CNN structures may affect performance on particular dataset, for example, in the research of V. Sunanthini, et.al., different model structure were shown to have different capacity for extracting features from fundus images [5]. Also, with different optimizers used for the step of updating parameters of neural network, which tends to be an important step for the learning process, the performance of CNN may also differ under the same task, as shown by Abbas Zohrevand and Zahra Imani in their research, VGG11 network tends to perform better on both MNIST, CIFAR10 and Fashion MNIST datasets when adopting Adam and AdaGrad optimizers than Stochastic Gradient Descent (SGD) and Adadelta [6].

Following the trend of discovering effects of different factors on the performance of CNNs, this article aims to compare the performance of ResNet-18 convolutional neural network with different optimizers when extracting different complexity of features (under different image classification datasets) with specific learning rates. In this article, 2 different datasets, Cats and Dogs dataset and Gender Classification datasets, were used to evaluate the performances, and 3 relevantly popular optimizers, Stochastic Gradient Descent, Adam optimizer and RMSprop optimizer, were employed for the experiment.

2. Method

2.1. Convolutional Neural Network (CNN)

As a representative algorithm component in the field of deep learning, CNN plays an important role in a large number of image classification algorithms. The basic principle of CNN is to extract features from complex visual inputs without human intervention, which exactly fits the goal of image classification tasks [7]. A complete CNN usually contains several convolutional layers, pooling layers and fully connected layers. Also, loss functions and optimizers are important for the back propagation part for the neural network to learn new parameters, further detail about this part will be discussed in later sections.

The convolutional layers, which is usually composed of several convolution kernels of certain size, with strides and padding techniques to adjust image size and density of convolving, can transform complex visual space structure and correlations between pixels into feature maps [7]. To filter important features from the feature map obtained by convolutional layers, the pooling layers, mainly max pooling layers or average pooling layers, are used. After several convolution and pooling steps, when the features of original input, which are representative and important enough, are extracted, these features will then be processed as input of Fully Connected (FC) layers for analysis.

There are different types of CNNs for solving image classification problems with different layer arrangements and inner structures, such as AlexNet, ResNet, GoogleNet, MobileNet, etc. In the next section, ResNet-18, the structure used for the experiment in this article, will be introduced.

2.2. Residual Neural Network (ResNet)

Among different types of neural network structures, residual neural network, which is also called ResNet, tends to reduce the training difficulty for deeper networks [4]. This ability of ResNet relies on a characteristic way of layer-wise connection called residual block. Within the scope of a residual block several layers are usually included, not only direct connections between layers are included, but also shortcuts connections between non-directly connected layers are included. The ResNet is composed of stacked residual blocks and due to the shortcut connections between layers, this structure is proved to effectively help solve the degradation problem, which is that when more layers added to an already suitably deep model, the accuracy will begin to degrade rapidly, with error rate also increase to a high level [4].

In this article, ResNet-18 is used because this model can effectively avoid degradation problem mentioned above, as a light-weight model among ResNet models, this model can save calculating power and time for the small datasets adopted by the experiments while maintaining high accuracy of predictions. In the research of Modupe Odusami, et.al., a fine-tuned ResNet-18 network was used for the detection of early stage features of Alzheimer's Disease, and the results showed that for most of the

DOI: 10.54254/2755-2721/46/20241177

multiple types of classification tasks related to different types of functional brain changes, the network had over 95 percent of accuracy, which showed the capacity and reliability of ResNet-18 network in image classification task [8].

2.3. Optimizers

The convolutional layers, pooling layers and FC layers enable CNN to have the ability to extract features from complex visual images. However, in the real image classification tasks, the CNN still needs to learn from previous experiences to adjust its parameters in order to obtain better accuracy. The loss function and optimizers provide the way for a CNN to evaluate its performance and to update its parameters for improvements. There are different kinds of optimizers, and for the experiment in this article, three commonly used optimizers, Stochastic Gradient Descent (SGD), Adam optimizer and RMSprop optimizer were used.

2.3.1. Stochastic Gradient Descent (SGD). SGD is further derived from the earliest optimization method Gradient Descent (GD). The basic principle of SGD is the same as GD, which is to construct the loss function based on the gradient, which is obtained based on the partial derivative of the model output with respect to each of the parameters of the model. This gradient value in loss function can be used to evaluate the influence of each parameter on the quality of model and model outputs, by keep updating model parameters based on the value of loss function through iteration, the loss value will converge to the optimal [9]. In this process, the model will learn based on the iterative updates of parameters, and its performance will likely be improved after learning.

The difference of SGD to traditional GD method is that in SGD only one or a small batch of data in the dataset is randomly chosen to update the gradient per iteration, while GD method often uses the entire dataset for the update step, which means that instead of calculating the entire gradient value, an unbiased estimate of the gradient value is calculated at each iteration [9]. In contrast with traditional GD, SGD is proved to be more efficient and has a better chance of reaching global optimum solution [9].

2.3.2. Adam. Adam optimizer, or Adam optimization method, also called adaptive moment estimation, is a type of stochastic optimization that is efficient requires little memory requirement, and usually provide more accuracy by computing individual adaptive learning rates of each parameter from the estimated values of first and second moments of the gradients [10]. By introducing the concepts of moment, which plays an important role to enhance the speed and stability of convergence, problems that may occur in the process of learning can be controlled. Also, with the bias correction at each iteration, this optimization method has the ability to solve the inaccuracy of estimations occurred in previous iterations.

As one of the most commonly used optimization methods, Adam optimizer is proved to have outstanding performance in complex image classification tasks. According to the research of potato leaf disease prediction conducted by Shikha Prasher, et.al., the CNN used for disease detection with Adam optimizer can reach an accuracy of 96 percent [11], which can prove the performance of this optimization method.

2.3.3. RMSprop. Similar to Adam optimizer, RMSprop optimizer provide each parameter of the model an adaptive learning rate, and update each parameter based on its own learning rate. Comparing to traditional gradient descent methods, which employ fixed learning rate for all parameters and often find themselves difficult to strike a balance between convergence speed and stability, especially in high-dimensional and non-convex optimization landscapes, methods such as RMSprop and Adam can obviously mitigate this issue and can therefore allow the algorithm to efficiently navigate complex optimization spaces and converge faster than traditional gradient descent variants.

What different from Adam optimizer is that RMSprop maintains a moving average of the squared gradients for each parameter, while Adam optimizer also maintains an estimate of moment. Also, from the perspective of hyperparameters, RMSprop often contains less hyperparameters than Adam optimizer,

DOI: 10.54254/2755-2721/46/20241177

hence requires less focus on parameter tunning in the training process. On datasets without many noises, RMSprop can have satisfactory converging behavior and stability, in the research of Shikha Prasher, et.al. with respect to potato leaf disease prediction, RMSprop tended to have best performance, reached the accuracy of 97.6 percent [11].

3. Result

3.1. Training Details

The experiment of this article mainly focused on the difference of performance of ResNet-18 on different dataset with SGD, Adam and RMSprop optimizer with specific value of learning rate. In the process of experiment, the Kaggle Cats and Dogs image dataset, which has relatively clearer features for CNN to extract, and the Kaggle Human Gender Classification dataset, which obviously contains more complex features, are chosen for training the neural network.

For the pre-processing step, random cropping and resizing input image to the size of 224×224, random horizontal flip and conversion to PyTorch tensor were implemented. For the model, the pretrained ResNet-18 network was used, and all parameters were unfrozen for adaptability.

During the process of experiment, for both Cats and Dogs and Gender Classification dataset, the model was run with SGD, Adam and RMSprop optimizers separately, with learning rate of 0.001 and 0.0005. Each scenario consisting the combination of dataset, optimizer and learning rate was run for 20 epochs, and the accuracy and loss of last epoch was recorded for comparation. Also, the change of accuracy and loss were plotted in order to present the process of converging and learning. After results with all combinations were obtained, analysis was performed and several patterns in the training process, and possibly related reasons were discussed, as can be seen later in the Discussion section.

3.2. Performance

First, for each configuration of dataset, learning rate, and optimizer, the accuracy and loss value of the ResNet-18 model was recorded at the last epoch of totally 20 epochs for comparison and further analysis. Table 1 and Table 2 show the accuracy and loss value of the ResNet-18 network under each dataset, learning rate and optimizer configuration.

Table 1. Performance comparison on Kaggle Cats and Dogs dataset.

Learning Rate		0.0005			0.001	
Optimizer	SGD	Adam	RMSprop	SGD	Adam	RMSprop
Accuracy	91.46%	91.91%	92.29%	92.35%	92.27%	91.83%
Loss	1.22%	1.18%	1.13%	1.11%	1.13%	1.20%

Table 2. Performance comparison on Kaggle Gender dataset.

Learning Rate		0.0005			0.001	
Optimizer	SGD	Adam	RMSprop	SGD	Adam	RMSprop
Accuracy	78.80%	78.71%	78.64%	78.76%	77.70%	77.96%
Loss	2.80%	2.82%	2.85%	2.79%	2.93%	2.94%

The RMSprop result on Cats and Dogs dataset, and Adam and RMSprop result on Gender Classification result under learning rate 0.001 was noticeable. With a higher learning rate 0.001, the accuracy seemed to reduce comparing to the results with learning rate 0.0005, which showed that as the consequence of this possibly plethoric learning rate, divergence may occur during the process of training, which hindered the model from converging. This pattern also happened in experiments on both Cats and Dogs and Gender Classification datasets, and it seemed that the extent of the occurrence of this problem was different for the two datasets. To further investigate this pattern, the accuracy of each epoch in the training process was recorded and plotted, as shown in Figure 1 and Figure 2.

DOI: 10.54254/2755-2721/46/20241177

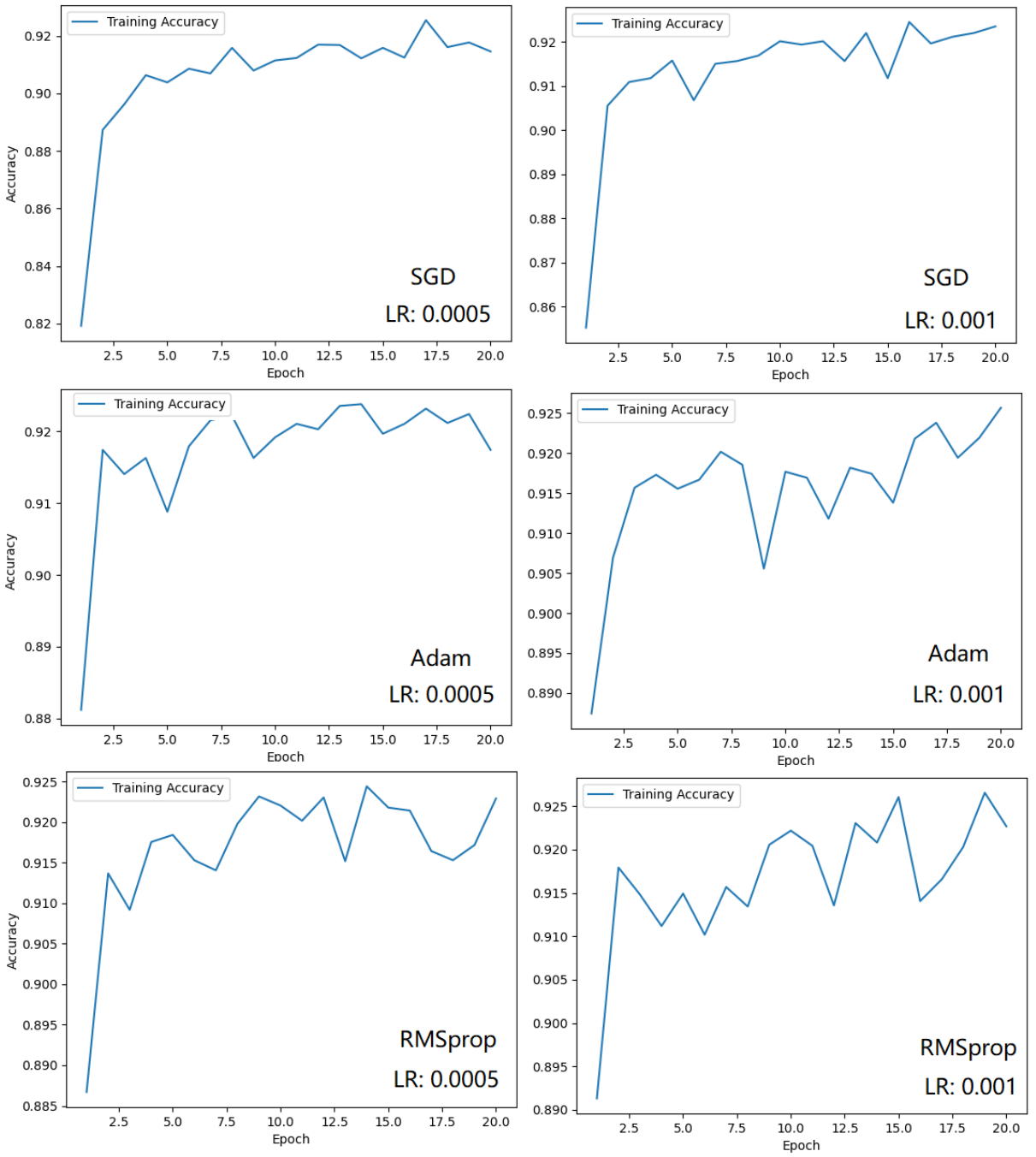


Figure 1. Accuracy curve on Cats and Dogs classification dataset (Figure Credits: Original).

4. Discussion

What can be attributed from the results table is that for the network, both the accuracy and loss value under every configuration on Cats and Dogs dataset is apparently higher than the accuracy and loss under Gender Classification dataset. Since Cats and Dogs images contains less details and complexity when comparing with Human Gender images, this can show that the complexity of inputs of CNN model can have a relatively large extent of influence with respect to the accuracy of the model, with more complexity, the accuracy of model classification seems to reduce.

DOI: 10.54254/2755-2721/46/20241177

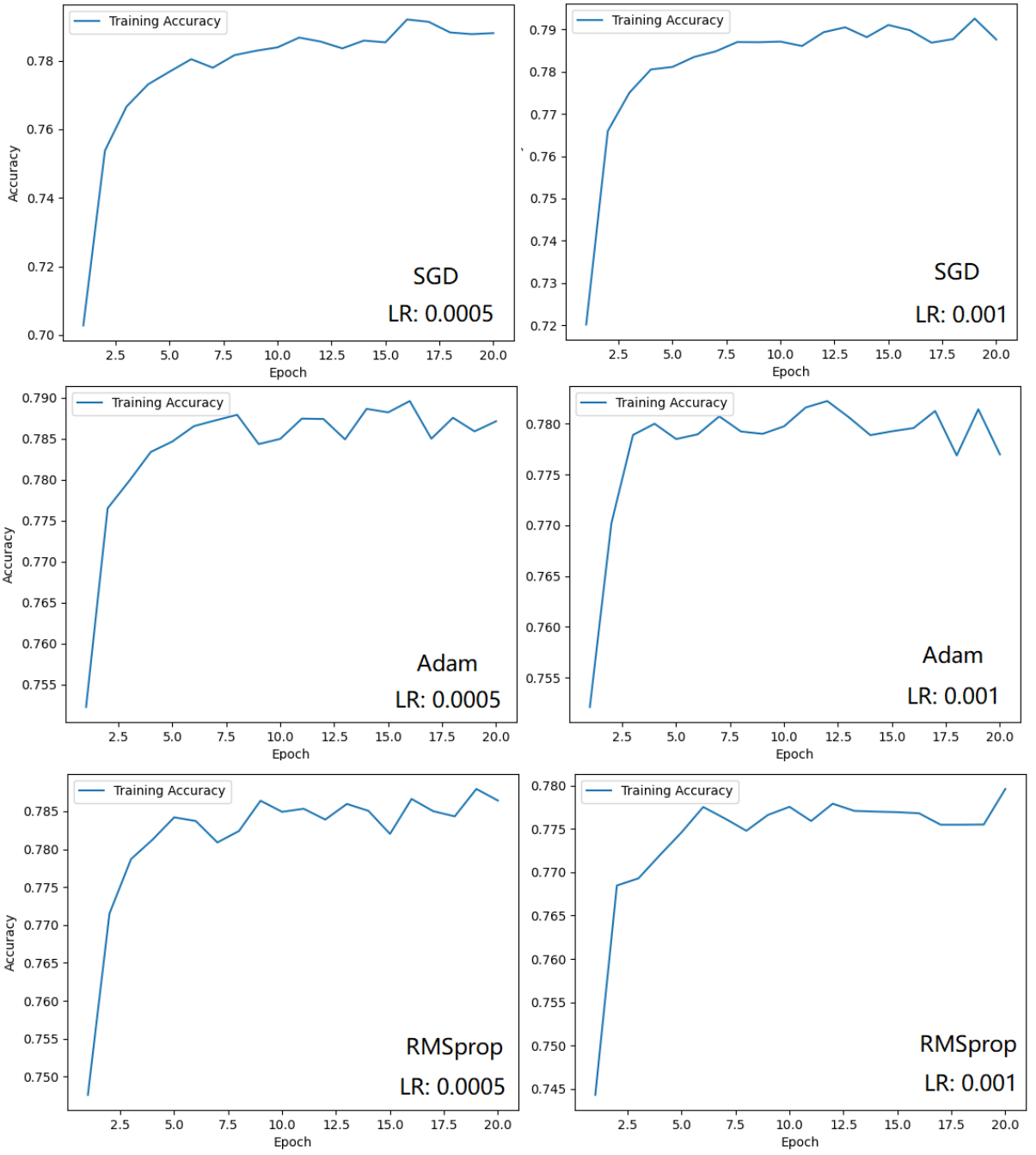


Figure 2. Accuracy curve on Gender classification dataset (Figure Credits: Original).

From the plots it could be noticed that there was relatively large extent of oscillations in the training process of Adam and RMSprop optimizer on Cats and Dogs images both when learning rate is 0.001, with lower learning rate was adopted, this oscillation was mitigated for some extent. Also, for experiments on Human Gender Classification, the trend of oscillation in the training process of Adam and RMSprop optimizers was much smaller than experiments on Cats and Dogs dataset, which may indicate that the complexity of input can also affect the speed and stability of convergence.

DOI: 10.54254/2755-2721/46/20241177

However, since Adam and RMSprop optimizer are pointed to have more sensitivity to the base learning rate and linear combination parameter of second-order moments, and one of the possible reasons of the divergence of Adam and RMSprop is the incorrect parameter settings [12]. In order to deeply investigate the patterns observed above, further experiments should be conducted to investigate for the possibility of inappropriate parameter settings which triggered the oscillation, or divergence behavior during the training process. Extra works with more learning rate options and more specific parameter tunning should also be performed in order to observe the pattern in a more general scope.

5. Conclusion

In this article, experiments of image classification using ResNet-18 under different configurations of learning rate and optimizers on two different image datasets with different level of complexity was conducted, and the experiment results were compared and analyzed. The experiment used SGD, Adam and RMSprop as optimizers, and two options of learning rate, 0.0005 and 0.001 were provided. As observed from the experiment results, performance oscillations, which may be attributed to plethoric learning rates, appeared in the process of model training, and differed as the model was trained on different data complexity (Cats and Dogs images and Gender Classification images). This can probably show that the complexity of data input can have influences on not only the accuracy of classification, but also on the convergence behavior, such as converging speed and stability.

References

[1] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... & Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164.

[2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1-9.

[3] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

[5] Sunanthini, V., Deny, J., Govinda Kumar, E., Vairaprakash, S., Govindan, P., Sudha, S., et al. (2022). Comparison of CNN algorithms for feature extraction on fundus images to detect glaucoma. Journal of Healthcare Engineering, 2022. 1-10.

[6] Zohrevand, A., & Imani, Z. (2022). An empirical study of the performance of different optimizers in the deep neural networks. In 2022 International Conference on Machine Vision and Image Processing, 1-5.

[7] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 1-21.

[8] Odusami, M., Maskeliūnas, R., Damaševičius, R., & Krilavičius, T. (2021). Analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network. Diagnostics, 11(6), 1071.

[9] Sun, S., Cao, Z., Zhu, H., & Zhao, J. (2019). A survey of optimization methods from a machine learning perspective. IEEE transactions on cybernetics, 50(8), 3668-3681.

[10] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[11] Prasher, S., Nelson, L., & Jagdish, M. (2023). Potato leaf disease prediction using RMSProp, Adam and SGD optimizers. In 2023 International Conference on Advancement in Computation & Computer Technologies, 343-347.

[12] Zou, F., Shen, L., Jie, Z., Zhang, W., & Liu, W. (2019). A sufficient condition for convergences of adam and rmsprop. In Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, 11127-11135.