Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/35/20230401

Automated machine learning-based neural network for brain tumor classification

Shim Ho

Department of physics, Hong Kong University of Science and Technology, Hong Kong SAR, 999077, China

hshimaa@connect.ust.hk

Abstract. In the contemporary medical landscape, characterized by the widespread and pernicious affliction of brain tumors, the imperative to enhance diagnostic modalities is paramount, aligning with the overarching objective of precisely delineating the affected patient cohort, thereby affording the prospect of administering expeditious therapeutic interventions. Within this scholarly discourse, the principal objective pertains to the discernment of an optimal Convolutional Neural Network (CNN) model, engendered through the mechanizations of automated machine learning, as instantiated within the virtual precincts of the "Edge Impulse" online platform. The corpus of investigation entails the acquisition of requisite data sets from the digital repository denominated "Kaggle," specialized in the domain of scientific knowledge. The amassed data sets, having undergone meticulous preprocessing procedures, were subsequently subjected to partitioning activities within the confines of the "Edge Impulse" framework, wherein a standardized ratio of division, namely a four-to-one proportionality between training and testing subsets, was consistently maintained across discrete data clusters. The training and testing processes were accomplished on Edge Impulse. The image mode, data learning block, learning rate, et cetera, were modified for each neural network models trained on Edge Impulse. Each model performed differently, and the distinct testing accuracy and on device performances were collected for each model for comparison. The experimental results demonstrate that using transfer learning supported by Edge Impulse with learning rate equals to 0.00051 and “fit longest axis” image resize mode is the optimal option for training a brain classifying model on Edge Impulse through automated machine learning.

Keywords: Convolutional Neural Network, artificial intelligence, brain tumor, classification.

1. Introduction

The term "brain tumor" denotes an aberrant cellular development within the cerebral tissue, which may originate intracranially to form a primary tumor or result from metastatic propagation originating from extracranial sites. There are two general types of brain tumors: benign which is non-cancerous and malignant which is cancerous. Astrocytomas and glioblastomas may develop from glial cells in the brain or from the meninges to form meningiomas [1]. Evidently, brain tumors are increasingly prevalent, with at least one million Americans afflicted by primary brain tumors, and projections indicating a notable rise in new diagnoses, reaching approximately 94,390 by the conclusion of 2023. Furthermore, brain tumor is also one of the most harmful and lethal diseases in the world. The death rate of having a malignant brain tumor is 35.7% in 2023, which will approximately cause 18990 people die in America [2]. Apart from death, the living patients will suffer from numerous side effects caused by the tumors, such as learning and understanding difficulty, depression, fatigue, and most importantly, the personality changes [3]. It is confirmed that Artificial Intelligence (AI) with Convolutional Neural Network (CNN) can improve the efficiency of diagnosing brain tumor [4, 5]. For instance, through computation and complicated algorithms, the AI can return a deep and complex analysis to a patient's diagnosis without medical practitioners doing time-consuming processing on the Magnetic Resonance Imaging (MRI) data. Additionally, AI are capable at detecting distinct diseases including brain tumor at their early stages, potentially affording patients the advantage of detecting their malignancies during less advanced phases, consequently augmenting prospects for successful therapeutic outcomes [6]. This underscores the pivotal significance of advancing enhanced high-performance AI models to deliver heightened diagnostic precision, thereby extending tangible benefits to a wider array of patients.

Due to the consistent development of AI, AI is now utilized in various fields such as controlling power system in power industry to reduce cost errors and offer optimal energy planning. Also, it allows more frequent real-time control actions update to the industry through applying fuzzy logic algorithm [7,8]. It is hard to proceed in the topic of brain tumor detection and classification without the contribution of previous studies. Feature Attention Network (FAN) was developed to deal with category imbalance and inconsistency of the target object scale, it used top-down feature fusion mechanism to extract the crucial features. FAN could also learn the correlations of the extracted features and the dependencies of labels [9]. This network makes valuable improvement on multi-label classification neural networks' development, which helps to deal with problems like multi-label tumors classification. It is worth noticing that AI is also introduced for solving other medical issues Zaidan et al had used AI to classify COVID-19 medical images. Their model used Multi-Criteria Decision Analysis (MCDA) to evaluate the multi-complex attribute of the medical image and set up benchmarked AI techniques for COVID-19 medical image classification [10]. Utikal et al had trained CNN using ResNet 50 architecture for classifying over 300 different kinds of lesions which were verified by biopsy, and the accuracy of people with using CNN achieved 82.95% [11]. If the CNN model is implemented into clinics, it can increase the efficiency of diagnosis significantly.

Almost all the experiments and developments on AI mentioned previously in this paper used coding and commands from computer programming languages. However, in this research, Automated Machine Learning (AutoML) was used instead of the traditional coding and time-consuming methods for building AI. The brain tumor classifying CNNs were going to be trained with AutoML on an online platform called "Edge Impulse", with preprocessed MRI images selected from an online database called "Kaggle". The aim of this research was to produce the best possible performance CNN model on Edge Impulse platform and found out the reliability of CNNs that were trained with AutoML.

2. Methodology

2.1. Dataset acquisition

The data acquisition was implemented on one of the most comprehensive scientific databases called "Kaggle". By conducting a time-consuming selection manually, a curated and preprocessed dataset with obvious tumor features in each image was chosen. The size of the dataset was 91MB and the dataset contain 2 sets which were training and testing set. Within each set, there were 4 distinct classes, namely Glioma tumor, Meningioma tumor, No tumor, Pituitary tumor.

Before training a CNN model on Edge Impulse, the selected dataset had to be imported. Data import was achieved in "Data acquisition" block which allowed to upload the downloaded local dataset onto the online platform. The selected dataset with 3, 041 images in jpg form was uploaded in "Data acquisition" block. During the process of uploading, all images were labelled and separated into four classes automatically by recognizing the name of the uploaded file that they were in. As stated previously that the images had distributed into training and testing sets already by the uploader. Nonetheless, the initial distribution within each class were different between classes. For instance, meningioma tumor class had 210:717 testing to training set ratio which testing set was approximately 22.7% and training set constituted the other 77.3%. But no tumor had 51:304 testing to training set ratio which testing set contained 14.4% of total and training had 85.6%. To ensure every class had approximately the same training to testing set ratio, the images in each class were redistributed by Edge Impulse to get a 2:8 testing to training set ratio within each class shown in Table 1. This division allowed models to learn features and associations to the greatest extend during supervised learning, while also giving models sizable amount of information to assess how well they performed during the training process [12].

Table 1. Data structure.

Classes:	Glioma tumor	Meningioma tumor	No tumor	Pituitary tumor
Training patch size:	725	744	283	683
Testing patch size:	181	183	71	171

2.2. Edge Impulse-based recognition

The comprehensive training and testing procedures were conducted within the purview of the web-based platform named "Edge Impulse." The cutting-edge platform Edge Impulse allows for the creation and implementation of embedded machine learning models. It offers an intuitive interface for categorizing, gathering, and processing sensor data as well as for training and deploying machine learning models directly on edge devices, enabling effective and real-time AI applications. The platform also has numerous advanced and convenient features. For instance, it supports live classification, which is a function allowing another device to deal with a related real-life problem by using the trained model on Edge Impulse. The platform also has versioning function to store all the data, settings, preliminary findings, and finished models. By adding a revision to a new project, the model can return to earlier iterations of design easily.

The block “Impulse design” on Edge Impulse allowed users to build and train structures of different CNN models. It contained three sub-blocks which were “Create impulse”, “Image”, and “Classifying” with distinct functions. The overall structure of a CNN model was modified in “Create impulse”. In this sub-block, various data processor for feature extraction and learning blocks for classifying data could be chosen to find out the combination of best performance. Because the images from the dataset were preprocessed, hence, image processor supported by Edge Impulse was selected to be the data processor (the other option left was raw data processor supported by edge impulse). There were two reliable learning blocks which were transfer learning and classification supported by Edge Impulse and they were used to be trained and compared between each other. The other learning blocks were specialized in object detection, movement recognition, audio classification, and finding outliers by K-means, which were all non-related with the research topic. There was another possible learning block option, which was transfer learning (image) supported by BrainChip Akida™. Nevertheless, it only supported AKD1000 MINI PCIe board, so it was excluded by hardware reasons. Additionally, input pixels and size mode between different models were set in this sub-block. Each model was trained using three different size mode which were “fit longest axis”, “fit shortest axis”, and “squash” to find out the most suitable mode which generated the highest accuracy. Notice that all experiments were proceeded in 96x96 pixels input image. All models were set to have RGB color depth in “Image” sub-block. In addition, the features of every image were generated in “Image” sub-block for later training process as shown in Figure 1.

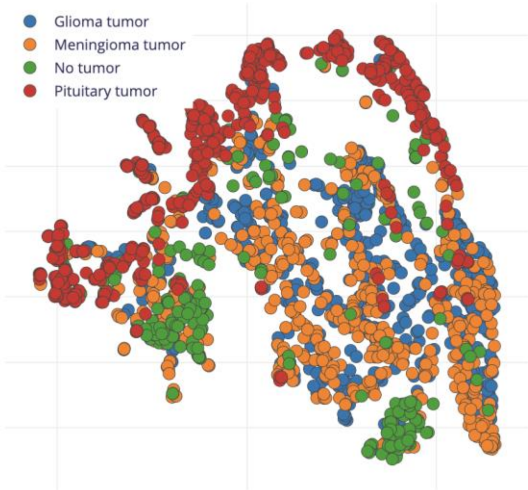

Figure 1. Features generated by Edge Impulse (Photo/Picture credit: Original).

The neural models were subjected to training within the "Classifying" sub-block, preceded by adjustments in the number of training cycles (epochs), learning rate, and the underlying neural network architecture. The training cycles for all models were set to be 20 epochs. To define a proper initial learning rate for different models, the models were tested from learning rate of 0.000005 to 0.05 with increasing scale of 10 each time, the one which returned the highest testing accuracy was chosen to be the initial learning rate. The learning rate for each model was changed gradually by 1% increment of its initial learning rate (with other variables fixed) in each successive training process to find out the optimal learning rate that produced global maximum accuracy. The data augmentation option was kept closed during the whole experiment, which controlled this variable option for comparison between different models and reduce the training time. It was particularly noticeable that all models were targeting Arduino Portenta H7 (Cortex-M7 480MHz) during their training process. Moreover, the transfer learning model was using MobileNetV2 96x96 0.35 (final layer: 16 neurons, 0.1 dropout) as its neural network architecture. The classifier model used two 2D convolutional layers and two pooling layers, both 2D convolutional layers got kernel size equaled to 3. Additionally, one 2D layer had 32 filters and the other one had 16 filters. The training set was inputted to the networks when the settings in this sub-block were confirmed. After completion of training, the CNNs outputted training accuracy, lose rate, and confusion matrix. Furthermore, the trained CNN models were tested in “Model testing” block through classifying the testing set prepared in data organization step, the training accuracies were the outputs.

3. Results and discussion

3.1. The performance of the model

According to Figure 2, by increasing the learning rate from 0.000005 to 0.05 with scale of 10 with epoch that equals to 20 and batch size that equalled to 3041, the model using transfer learning had the highest testing accuracy which was 79.21% when learning rate equated to 0.0005. The image size mode adopted by this model was denoted as "fixed shortest axis."

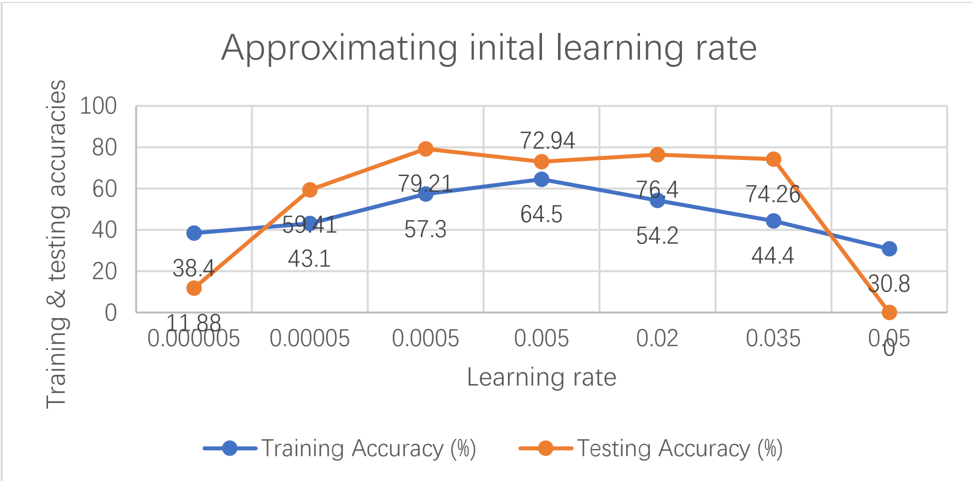

Figure 2. Approximating initial learning rate (Photo/Picture credit: Original).

Model 1 employed transfer learning supported by Edge Impulse as its learning block, its epoch that equalled to 20, batch size that equalled to 3041 and using “fit shortest axis” mode. According to Figure 3, model 1 achieved the highest testing accuracy of 81.19% when learning rate equalled to 0.00051.

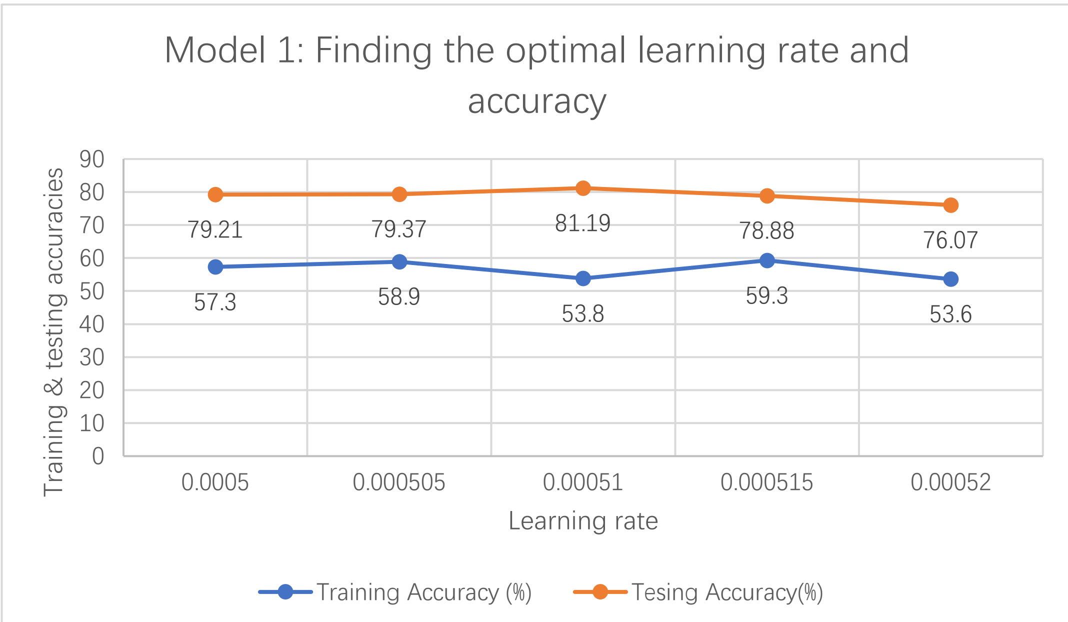

Figure 3. Model 1 performance with varying learning rate (Photo/Picture credit: Original).

Model 2 used classifier supported by Edge Impulse as its learning block, its epoch was 20, batch size equalled to 3041 and using “fit shortest axis” mode. According to Figure 4, model 2 got the highest testing accuracy of 76.57% when learning rate equalled to 0.00053.

DOI: 10.54254/2755-2721/35/20230401

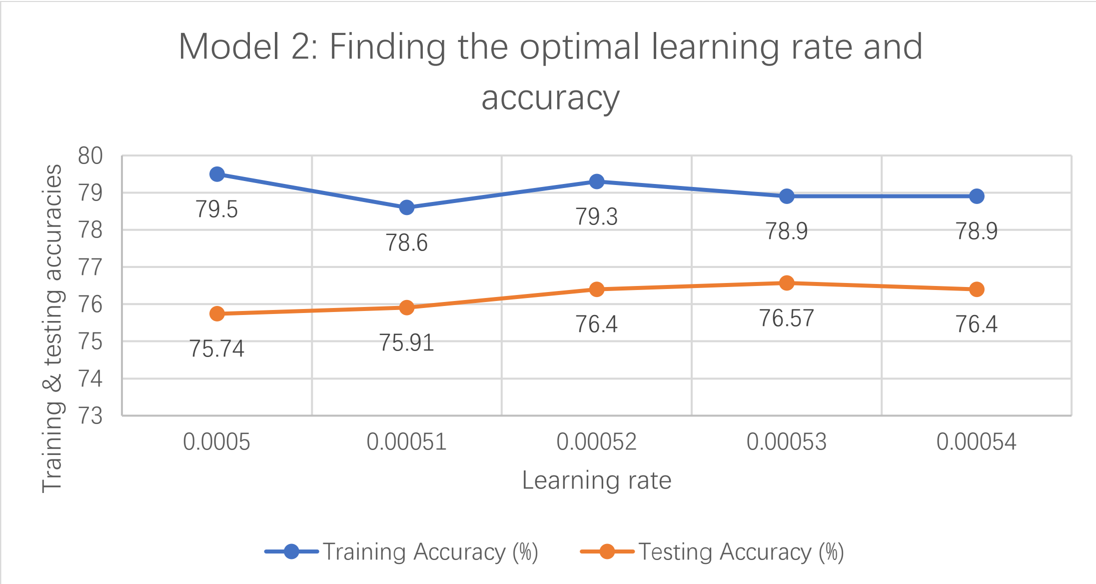

Figure 4. Model 2 performance with varying learning rate (Photo/Picture credit: Original).

From Figure 5, it was evident that model 1 obtained the highest testing accuracy which was 80.86% when the size mode of image is "fit longest axis".

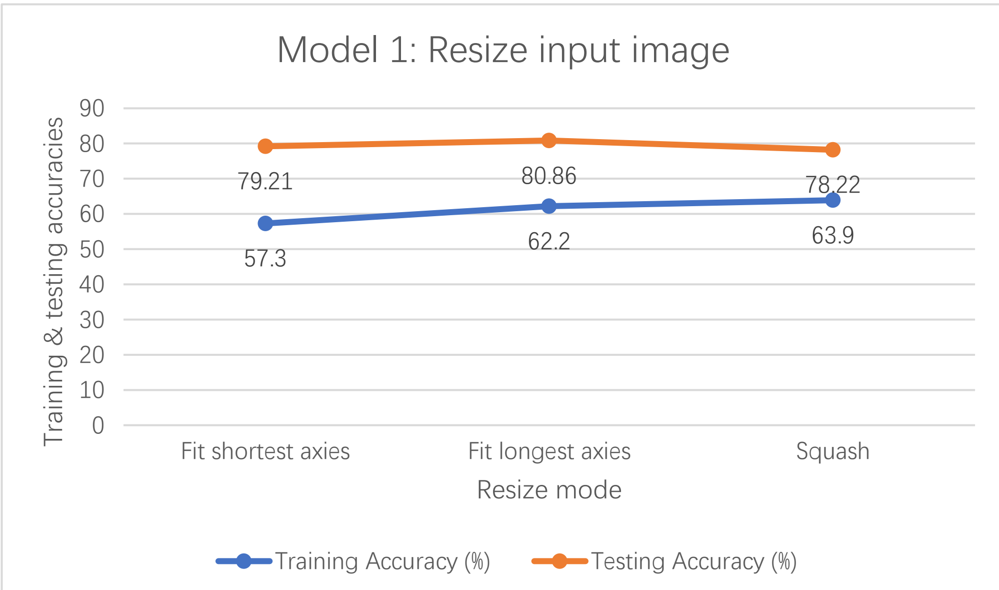

Figure 5. Model 1 performance with different size mode (Photo/Picture credit: Original).

Table 2 demonstrated the time consumed by model 1 for processing a new data and the computer memory usages. Model 1 was using "fit shortest axis" image mode during this test.

Table 2. On device performance of model 1.

Model 1 (RGB)	Fit shortest axis	
Inferencing time:	Peak RAM usage:	Flash usage:
108ms	333.8K	579.9K

Table 3 demonstrated the time consumed by model 2 for processing a new data and the computer memory usages. Model 2 was using “fit shortest axis” image mode during this test.

Table 3. On device performance of model 2.

Model 2 (RGB)	Fit shortest axis	
Inferencing time:	Peak RAM usage:	Flash usage:
2215ms	363.3K	66.7K

3.2. Discussion of experimental results

It is obvious that model 1 has higher best testing accuracy (81.19%) than model 2 (76.57%) with the same settings. According to Table 2 and Table 3, it is evident that model 2 gets roughly 20.51 times larger inference time than model 1. Model 2 also gets 29.5K extra peak RAM usage than model 1. Model 1 has approximately 8.69 times larger flash usage than model 2. Overall, model 1 gets 2 indexes (inference time and peak RAM usage) of on device performance better than model 2. Hence, from the perspectives of accuracy and on device performances, model 1 is the best model and the upcoming tests only focusing on model 1. Therefore, transfer learning supported by Edge Impulse is the best learning block for training a classification CNN model on Edge Impulse platform. This may because that Edge Impulse platform has used specialized data and a sizeable batch of data to implement the pre-train process of their transfer learning model, so that the convolutional layers have much better distribution of weights in different neurons in each convolutional layer. From Figure 5, “fit longest axis” mode produces the highest testing accuracy, it may because that it is the only mode which keeps all the original information. For “fit shortest axis”, the information on the edge of the input images is cut off when the computer tries to fit with the shortest axis with the least information. For “squash”, the original pixels were stretched so the RGB values for each pixel merge together, which cause confusion to the CNN and make it less effective on learning the feature of tumor. Therefore, “fit longest axis” is the best resize mode for training CNN model on Edge Impulse.

4. Conclusion

In this experiment, automated machine learning was proposed to train brain tumor recognition by using Edge Impulse online platform and a diversified dataset to avoid bias. The highest accuracy and best on device performance model on Edge Impulse was developed. Multiple experiments were implemented to find out the desired model. The experimental results indicated that transfer learning supported by Edge Impulse produced the best performance model from all perspectives. Additionally, “fit longest axis” was demonstrated to be the most desirable size mode for training image classification model. A limitation inherent to this study pertains to hardware constraints that precluded the utilization of certain learning blocks, resulting in their exclusion from the experimental paradigm. Subsequent endeavors could involve the prospective deployment of the extant model within clinical or hospital settings for real-time testing and learning applications.

References

[1] American Brain Tumor Association (n.d.) About Brain Tumors Availabe at: https://www.abta.org/about-brain-tumors/ (Accessed: 12 August 2023)

[2] National Brain Tumor Society 2023 Brain tumor facts Available at: https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-facts/#:~:text=35.7%25%20Relative%20Survival%20Rate%20for%20all%20patients%20with,die%20from%20a%20malignant%20brain%20tumor%20in%202023 (Accessed: 12 August 2023)

[3] Lupton A Abu-Suwa H Bolton G C and Golden C 2020 The implications of brain tumors on aggressive behavior and suicidality: a review Aggression and violent behavior 54 p 101416

[4] Qiu Y Chang C S Yan J L et al 2019 Semantic segmentation of intracranial hemorrhages in head

CT scans 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) IEEE 112-115
[5] Ayadi W Elhamzi W Charfi I et al 2021 Deep CNN for brain tumor classification Neural processing letters 53: 671-700
[6] Gull S and Akbar S 2021 Artificial intelligence in brain tumor detection through MRI Scans Artif Intelligence Internet Things pp 241-276
[7] Dahhaghchi I Christie R D Rosenwald G W and Liu C C 1997 AI application areas in power systems IEEE expert 12(1) pp 58-66
[8] Ahmad T Zhu H Zhang D Tariq R Bassam A Ullah F AlGhamdi A S and Alshamrani S S 2022 Energetics Systems and artificial intelligence: Applications of industry 4.0 Energy Reports 8 pp 334-361
[9] Yan Z Liu W Wen S and Yang Y 2019 Multi-label image classification by feature attention network Ieee Access 7 pp 98005-98013
[10] Albahri O S et al 2020 Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: Taxonomy analysis, challenges, future solutions and methodological aspects Journal of infection and public health 13(10) pp 1381-1396
[11] Hekler A et al 2019 Superior skin cancer classification by the combination of human and artificial intelligence European Journal of Cancer 120 pp 114-121
[12] Géron A 2022 Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow O'Reilly Media, Inc. chapter 2 pp 66-91