Research seminar

Analysis of the impact of convolutional neural network parameters on classification accuracy in a small medical image dataset

Author: Žan Peternelj, 89212061
Mentor: Peter Rogelj, PhD

University of Primorska
Faculty of Mathematics, Natural Sciences and Information Technologies

July, 2022

Abstract

In this work, we analyse the impact of individual parameters within a convolutional neural network (CNN) architecture with a small training data set on classification accuracy. We compared the influence of the following parameters: number of convolutional layers, size of convolution kernel, and number of filters. Classification was performed on 3D MRI medical images of brain lesions. Different neural network configurations were classified for a given lesion type against other types. This was done to investigate whether the same configurations are effective differently depending on the lesion type classification. In total 8 different CNN configurations were used to classify lesion types into two classes.

Key words
brain lesions, neural network, performance

1 Introduction

This study is an extension of our collaborative research with the University of Basel, where we have developed a neural network to classify brain lesions into two classes from five lesion types. This was done according to the impact of the presence of the certain lesion type on the course of the patient's disease. The data and their augmentation used in the study are taken from the aforementioned previous research. Given that the amount of training data is often limited, especially in the field of medicine, where the data is also obtained through the diagnostic process of the treatment itself, we investigated the influence of the parameters within the convolutional neural network on the performance of classification with the aim of designing an optimal architecture. The baseline architecture used is designed according to the results of the previous research, where simpler architectures proved to be more successful than more complex, deeper architectures. Previous studies [1, 2] have already found that the parameters of convolutional neural networks have an effect on classification accuracy on the same dataset. We focused our study on comparing the accuracy of different combinations of classifiers within a small dataset.

2 Methodology

2.1 Convolutional neural network architecture

The baseline neural network architecture used, represents the basic form of a convolutional neural network with one input layer, a convolutional layer and an output layer with two classification classes. Reason for selecting a simple architecture is that in previous research we achieved better results using such architectures. The following parameters were then tuned to the baseline architecture: number of convolutional layers, size of convolution kernels, and number of filters. In total, eight different architectures were used to compare the classification performance across each lesion type. For the parameter number of convolution layers, we added two and three convolution layer variants to the original architecture with one layer. For the size of the kernel parameter, we used two additional options, where in first case we reduced the original size from [5, 5, 5] to [3, 3, 3], and in the second case we increased it to [7, 7, 7]. We did the same for the number of filters, reducing the original number of 10 to 5 and increasing it to 20 and 30.

2.2 Data

The input data are 3D images of brain, in the context of the diagnosis of multiple sclerosis with classification of brain lesions. The lesions are classified into 5 different types [3], four of which can be attributed to a positive or negative effect on the development of the disease. Types two and six represent a favourable effect on the course of the disease, meaning that the patient's condition improves, while the presence of types three and five represent a negative effect. Type four cannot be attributed to an effect. As usual in medical image analysis, the amount of available data is limited, which has an impact on the development of the neural network architecture itself, as well as on the pre-processing of the data. The dataset consists of 5269 lesions, of which 1605 lesion cases are classified, representing only 30%. The remaining examples (lesions of type 1) were discarded as, for various reasons, they cannot be classified within the five types mentioned above, which would have a negative impact on the quality of the neural network learning. The number of cases per type is not evenly distributed, which also affects the learning ability. A more detailed breakdown of the number of cases per type is shown in the table 1. Due to the limited dataset, offline augmentation was performed to increase the amount of data. We used transformations with rotation along an arbitrary axis and mirroring. Before performing the augmentation, we split the dataset into training and test data in the ratio of 70% and 30%. By applying the augmentation, we increased the number of training data to a total of 58896 examples. A single lesion case is represented by a structure with four dimensions, where the first three represent the dimensions of the 3D image and the fourth and final dimension represents the channel. There are three channels, each for its own image of the same lesion area. First two images consist of two different MRI techniques, and the third images represent a mask indicating the area of the image where the lesion is located, as can be seen on figure 1. The lesion patches are 35x35x35 voxels in size.

Lesion type	Number of cases
1	3664
2	460
3	214
4	19
5	841
6	71
Sum	5269

Table 1: Number of lesions cases per lesion type.

2.3 Network training

Matlab software, version R2022a, was used to develop the neural networks and to train them. The neural network training was performed on a PC running Windows 11, Intel I9-12900K, 3.2 Mhz processor with 16 cores, Nvidia RTX3080 graphics card and 32 GB of working memory.

We used the same augmented training dataset and the same training settings to train all neural networks shown in table 4. Given that individual lesions within the training data are represented by different numbers of examples, we performed an equalization of the number of examples to the maximum possible number, or in other words, to the lesion type with the smallest number of examples, prior to learning. For lesion types 2, 3, 5 and 6, we performed learning on all architectures by classifying them into the

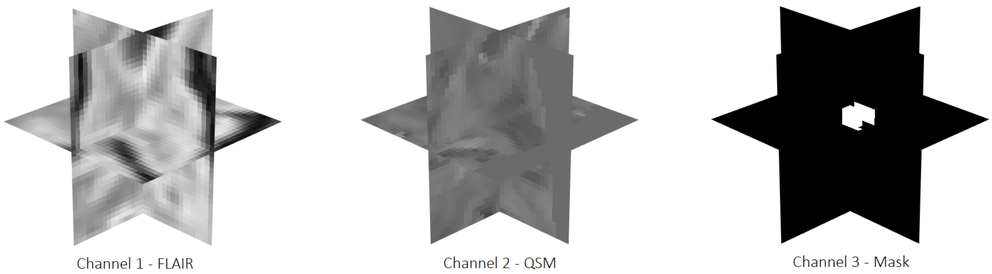

Figure 1: Example of lesion case. The size of the image is 35x35x35 voxels and it is has three channels, stored in the fourth dimension of the image. Each of the channel consists of a different image of the same lesion. First channel is an FLAIR MRI image, second channels is an QSM MRI image and the last channel is a mask of lesion location.

#	Lesion type	Definition
1	/	Lesion can not be classified.
2	Isointense	A lesion, but inside and around the lesion the values are the same.
3	Hyperintense rim	The border of the lesion has higher values compared to inside and outside the lesion
4	Hypointense rim	The border of the lesion has lower values compared to inside and outside the lesion.
5	Hyperintense lesion	The value inside the lesion has higher values compared to the surrounding, no specific delineation of the border is present
6	Hypointense lesion	The value inside the lesion has lower values compared to the surrounding, no specific delineation of the border is present

Table 2: Lesion type classification [3]

next two classes, the selected lesion type against all other lesion types. This resulted in 32 different models over which we performed testing with the original, not transformed lesion examples. We used two different datasets for testing, for the first one, similarly to the training data, we equalized the number of examples within the two classes, but not with augmentation, but just decreasing the number of lesion cases, while for the second one we kept the imbalance of the number of examples between the two classes.

#	Layer	Parameter
1	3-D Image Input	[35x35x35] x 3
2	Convolution	[5x5x5] x 10
3	Batch Normalization	
4	ReLU	
5	3-D Max Pooling	[2x2x2]
6	Fully Connected	
7	Softmax	
8	Classification Output	

Table 3: Baseline architecture

Option	Value
Initial learn rate	1e-3
Max Epochs	10
Shuffle	every-epoch
Learn Rate Schedule	piecewise
Execution Environment	GPU
Mini Batch Size	50
L2 Regularization	0.0005

Table 4: Train options used in the training for all architectures. The remaining options that not listed in the table used the default value.

3 Results

The results show that the accuracy of the classification of each lesion type is influenced by the architecture of the neural network. The results show that the parameters of the neural network layers affect the overall classification accuracy, but also that there are differences in classification performance between lesion types for the same architecture. There is also a difference between the two test datasets used, where the matched dataset achieves on average 3.9% lower classification accuracy, as can be seen in the Figure 2.

The architecture that achieved the highest results on average on the first test dataset is `filter 30`, while the lowest results were achieved by `conv[7,7,7]`. On the second data set, the best performing architecture was `2 conv`, and the worst was `filter 5`. The lesion type with the highest classification accuracy in both test datasets was type 6, with an accuracy of 91.3% in the first set and 86.4% in the second set. The worst performing type was type 5, with 74.6% on the first set and 72.4% on the second. The classification accuracy of each neural network architecture by lesion type on both test datasets is shown in the figures 3,4 On the first dataset, lesion type 2 was classified with the highest classification accuracy by architecture `2 conv` and `3 conv`, and with the lowest classification accuracy by architecture `conv[3,3,3]` and `filter 20`. On the second dataset, the highest classification accuracy was achieved by architecture `filter 30`, and the lowest by architecture `filter 20`. Lesion type 3 was classified with the highest classification accuracy by architecture `2 conv` and the lowest by architecture `conv[7,7,7]` on the first dataset. On the sec-

#	Architecture	Description
1	baseline	Baseline architecture.
2	conv[3,3,3]	Convolution kernel size decreased to [5,5,5].
3	conv[7,7,7]	Convolution kernel size increased to [7,7,7].
4	filter 5	Number of convolution kernels decreased to 5.
5	filter 20	Number of convolution kernels increased to 20.
6	filter 30	Number of convolution kernels increased to 30.
7	2 conv	Two convolution layers with same parameters as baseline.
8	3 conv	Three convolution layers with same parameters as baseline.

Table 5: Used CNN architectures and descriptions of changes to the parameters.

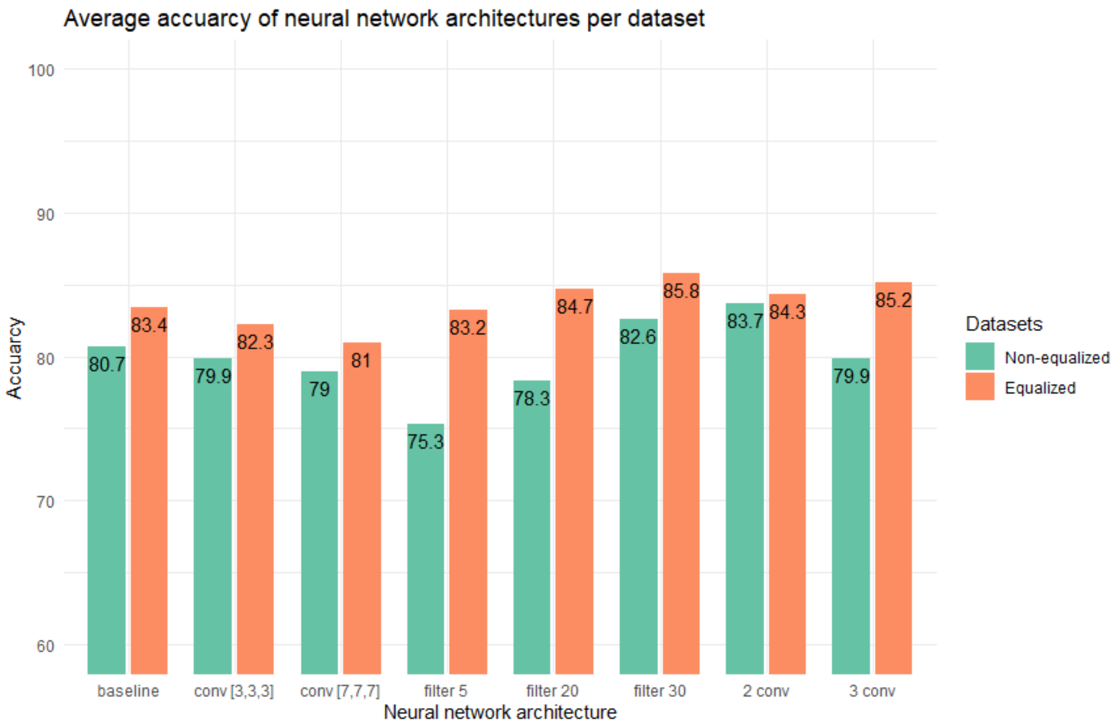

Figure 2: Classification accuracy by neural network architecture per lesion type.

ond test set, the highest accuracy was achieved by the `2 conv` architecture and the lowest accuracy by the `filter 30` architecture. On the first test set, the type 5 lesion was most accurately classified by the `filter 30` architecture and the lowest accuracy by the `filter 5` architecture. On the second test set the results were the same. For the lesion type 6, the highest classification accuracy on the first dataset was achieved by the `3 conv` architecture and the lowest by the baseline architecture. On the second dataset, the best performing architecture on type 6 was `2 conv`, while the lowest accuracy was achieved by `filter 5`.

4 Discussion

Effect of CNN layer parameters on classification accuracy using an unbalanced limited size dataset is analyzed in this paper. The results show, that configuring parameters achieves different results on different classifications, which is important to consider when optimizing the architecture for a given problem. The biggest improvement on average in comparison to baseline architecture was achieved by the `filter 30` on the equalized dataset, and by `2 conv` on the non-equalized datatset. Smaller kernel size `conv[7,7,7]` on average performed worse that baseline, except for the type 6 on equalized dataset and type 3 on the non-equalized dataset.

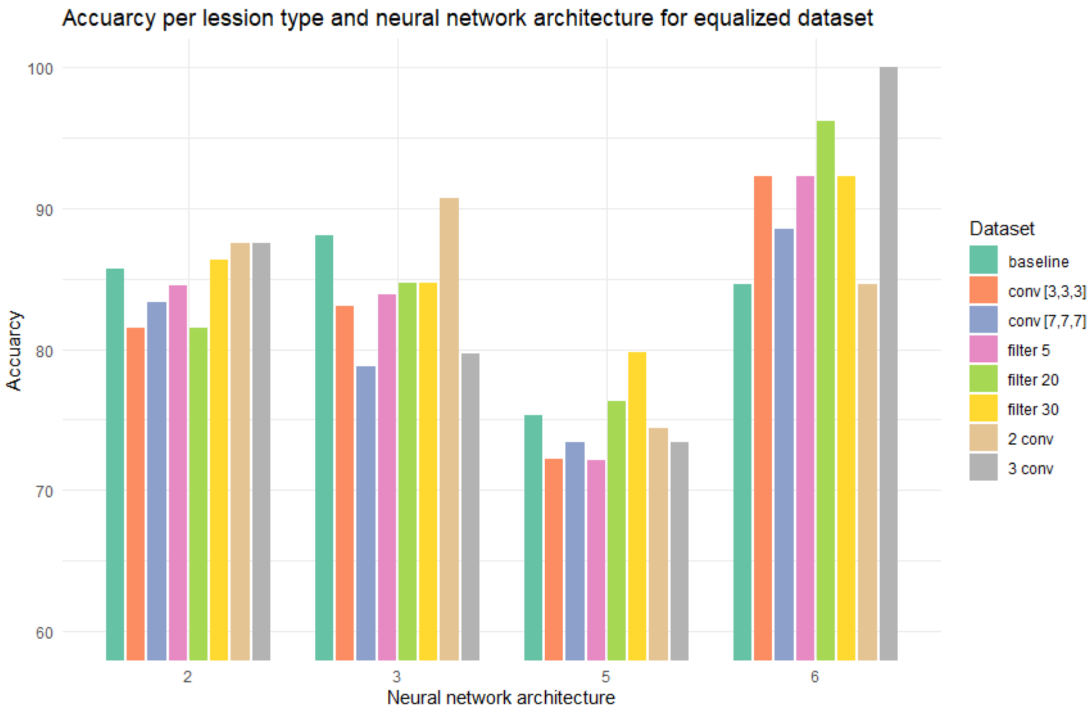

Figure 3: Classification accuracy per neural network architecture and lesion type for equalized dataset.

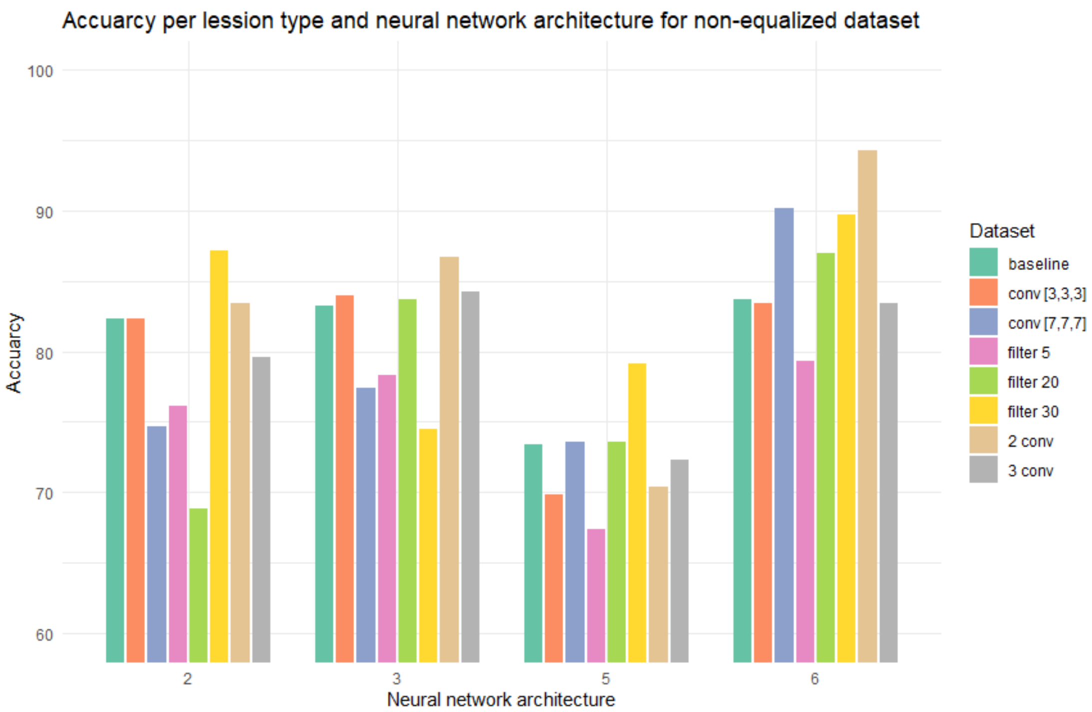

Figure 4: Classification accuracy per neural network architecture and lesion type for non-equalized dataset.

Bigger kernel size `conv[7,7,7]` showed overall decrease in acuraccy on both datasets, except for type 6 where its accuracy was higher compared to baseline. Decreasing number of filters `filter 5` achieved on average worse results than baseline architecure, except for type 6 in the equalized dataset. Overall the equalized dataset achieved higher accuracy, and this might be due to the smaller number of cases as a consequence of the equalization. Biggest difference between equalized and non-equalized dataset was found in `filter 5`. On the equalized dataset the highest accuarcy per type was achieved by `2 conv` and `3 conv` on type 2, `2 conv` on type 3, `filter 30` on type 5 and `3 conv` on type 6. On the non-equalized dataset the highest accuracy per type was achieved by `filter 30` on type 2, `2 conv` on type 3, `filter 30` on type 5 and `2 conv` on type 6. The results show that using a single architecture for all classifications would not achieve the optimal accuracy, as some types are classified more accurately by certain architectures than the others. Future research for this problem is needed, which would analyze the optimal way of combing different types of architectures used for each type classification, into one combined classification which would determine the final class.

References

[1] Sanjit Maitra, Rahul Kumar Ojha, and Kuntal Ghosh. Impact of convolutional neural network input parameters on classification performance. In *2018 4th International Conference for Convergence in Technology (I2CT)*, pages 1–5, 2018.

[2] Kaushik Gobindram Pasi and Sowmiyaraksha R. Naik. Effect of parameter variations on accuracy of convolutional neural network. In *2016 International Conference on Computing, Analytics and Security Trends (CAST)*, pages 398–403, 2016.

[3] Reza Rahmanzadeh, Riccardo Galbusera, Po-Jui Lu, Erik Bahn, Matthias Weigel, Muhamed Barakovic, Jonas Franz, Thanh D Nguyen, Pascal Spincemaille, Simona Schiavi, Alessandro Daducci, Francesco La Rosa, Martina Absinta, Pascal Sati, Meritxell Bach Cuadra, Ernst-Wilhelm Radue, David Leppert, Jens Kuhle, Ludwig Kappos, Wolfgang Brück, Daniel S Reich, Christine Stadelmann, Yi Wang, and Cristina Granziera. A new advanced MRI biomarker for remyelinated lesions in multiple sclerosis. *Ann. Neurol.*, June 2022.