Neural Network Based Classification of Brain Lesions

Nedim Šišić
Faculty of Mathematics, Natural Sciences and Information Technologies
University of Primorska
Koper, Slovenija
89202005@student.upr.si

***Abstract*— A brain lesion is an area of tissue injury or disease within the brain. Brain-imaging techniques, such as MRI, produce images of the brain used for diagnosis or prognosis of a wide variety of injuries or conditions. Such images can be used to classify brain lesions. Recently, neural networks have been used to great success in different computer vision domains, including brain image processing. In this work, we develop a convolutional neural network that automatically classifies MRI images in multiple sclerosis patients. The image dataset consists of volumetric (three-dimensional) images with three channels: FLAIR MRI, QSM MRI, and a mask channel indicating lesion location. We describe data augmentation techniques we use for increasing the dataset and present the network architecture and the results of the model. Specifically, we show the results of training the network using different combinations of the three image channels.**

Keywords—Computer vision, Neuroimaging, Convolutional neural networks, Multiple sclerosis

1. INTRODUCTION

Medical imaging produces images of the interior of a body in order to aid clinical analysis and medical intervention. Magnetic resonance imaging (MRI) is a technique that uses strong magnetic fields and radio waves to generate images of the organs in the body. MRI is widely used in hospitals and clinics for imaging the brain and assisting the diagnosis of neurological disorders. One such disorder is multiple sclerosis (MS). MS is a demyelinating disease, in which the insulating myelin covers of nerve cells in the brain and spinal cord are damaged, resulting in a range of symptoms. An estimated 2 million people worldwide currently have the disease (Reich et al., 2018). Recently, Quantitative Susceptibility Mapping (QSM) MRI has been shown to allow classification of MS lesions – areas of tissue abnormality – into demyelinated and remyelinated types (Rahmanzadeh et al., 2022). In remyelinated lesions, a repair process called remyelination attempts to create new myelin sheaths on the surface of demyelinated axons. Thus, demyelinated lesions indicate that the state of the patient is expected to worsen, while remyelinated lesions indicate the state is expected to improve. Classifying MS lesions into the two types based on QSM MRI may therefore be very beneficial in diagnosis of MS patients.

In general, medical images are usually manually processed by medical experts. This can prove to be a laborious and time consuming task, and so computer-aided systems are frequently used to reduce the time required for image processing and even improving diagnostic accuracy. Specifically, computer vision and image processing are used for automatic extraction of information from medical images; a review of such techniques can be found in (Gao et al., 2018, Yanase et al., 2019).

In the last decade, the field of computer vision has been strongly impacted by neural networks. Neural networks are machine learning models that can be applied to numerous problems, such as computer vision, natural language processing, playing games etc. A convolutional neural network (CNN) is a neural network architecture that has revolutionized many subfields of computer vision by giving state-of-the-art results in problems such as image classification (Al-Saffar et al., 2017), image segmentation (Minaee et al., 2020), and object detection (Zhao et al,. 2019). CNNs have been used in numerous medical imaging domains and have achieved remarkable results in various applications. However, medical imaging may often present challenges to the use of neural networks. Notably, neural networks are machine learning models that require large amounts of standardized data to be successfully trained – such datasets may not be available in many medical imaging tasks. In such cases, the number of sample images may be small, the labels associated with the images may be sparse, or the image categories may be heterogeneous and imbalanced. Thus, medical imaging might require additional efforts, commonly regarding data augmentation, when neural networks are to be used. A review of neural networks in medical imaging can be found in (Litjens et al., 2017).

In this work, we present a convolutional neural network for a novel problem in classification of multiple sclerosis MRI images. Each image is a volumetric (three-dimensional) image of size 35x35x35 and consisting of three channels: FLAIR MRI, QSM MRI, and a mask channel indicating lesion location. The

network classifies the lesions into demyelinated and remyelinated types. We show and compare the results of training the network using different combinations of the three image channels. Our results demonstrate that neural networks are a promising tool for this problem. Expanding the work may enable the use of a model such as ours in clinical practice, aiding medical experts in MS diagnosis by easing the burdensome task of image processing and improving diagnosis accuracy.

The paper is organized as follows. We give an overview of convolutional neural networks and their use in medical imaging in **Section 2**. We present the dataset used in our work and specify our goals in **Section 3**. In S**ection 4.1**, we describe the techniques we use for manipulating the dataset. We present the neural network architecture and the network learning parameters in in S**ection 4.2** and **Section 4.3** respectively. We give and discuss the results of training the network using different combinations of image channels in **Section 5**. Finally, we conclude the work and discuss opportunities for future work in **Section 6**.

2. PREVIOUS WORK

Before the mass popularization of neural networks in computer vision and image processing, problems in these fields were usually approached by manual engineering of image features or by less powerful machine learning models. Neural networks have since greatly overshadowed these traditional solutions. One of the most successful neural network architectures in computer vision and image processing is the convolutional neural network. The origins of the CNN can be traced to the "neocognitron" (Fukushima, 1980), a model introduced in 1980. However, it would not be until 2011 that CNNs would achieve state of the art results in image processing tasks (Dan Ciresan et al., 2011). Since then, CNNs have made a revolution and become a the standard approach to many problems in these domains.

A convolutional neural network is a specific kind of a neural network which keeps only the neuron connections that preserve spatial relationships in the data. By removing the rest of the connections, a CNN allows for faster, less memory demanding, and more certain learning of image features. Unlike a fully connected neural network, which consists of fully connected layers, a CNN uses *convolutional* layers, often interspersed with *pooling* layers[1].

Convolutional neural networks have found much use in medical imaging (Litjens et al., 2017). Specifically, a detailed survey of the use of CNNs in MRI image analysis is given by (Lundervold et al., 2018). A natural approach to using the previous successes of CNNs to our problem would be *transfer learning*, i.e., the transfer of network weights from a previously trained CNN. However, such weights are difficult to transfer between differing network architectures. The image dataset in our novel problem is very specific, consisting of volumetric images of size 35x35x35 and three image channels. We have not found networks that process such data and therefore had to train our network from scratch.

3. PROBLEM SPECIFICATION

3.1 Dataset

Our network classifies MRI images of multiple sclerosis lesions. Our lesion dataset was obtained as follows. MRI scanning was carried out on 115 MS patients (76 relapsing-remitting MS and 39 progressive MS), giving MRI images of different modalities. An automatic lesion segmentation procedure was performed on the FLAIR and MP2RAGE MRI images, resulting in masks indicating the lesion locations for each patient. The same procedure, conducted on a similar dataset, is described in (La Rosa et al., 2020). In total, 5250 lesions were registered (one patient may have numerous MS lesions).

For each lesion, a patch of size 35x35x35 voxels was created such that the lesion of interest is centered in the patch. In the remainder of the paper, we use "image" when referring to these patches as well. Each such image in our dataset consists of three channels, each channel being a 35x35x35 volumetric image:

- an MRI fluid-attenuated inversion recovery (FLAIR) image
- an MRI quantitative susceptibility mapping (QSM) image
- a mask image specifying the location of the lesion

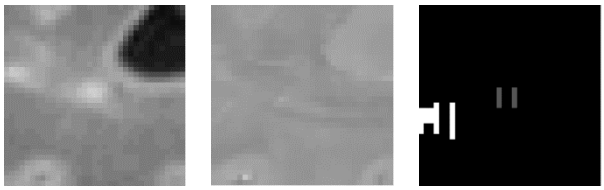

Figure 1: Three channels of a slice of an instance in the dataset: FLAIR, QSM, and mask.

We interpret the dataset images in the following way: each image has three spatial dimensions and one additional dimension consisting of the three channels. For that reason, we choose a 3D convolutional neural network as our neural network architecture.

The image dataset is split into 6 classes, according to image lesion type: **1** – unclassified lesions, **2** – isointense lesions, **3** – PRL lesions, **4** - hypointense rim lesions, **5** – hyperintense lesions, **6** – hypointense lesions. The

[1] For a detailed explanation of the convolutional neural network architecture, see the *CS231n Convolutional Neural Networks for Visual Recognition* Stanford course.

classification was originally performed manually by medical doctors using only the QSM MRI images in the dataset. In classes **2** to **6**, the lesions in different classes have different visual characteristics and carry different implications for the patient's diagnosis. The lesions in class **1** have not been able to be classified into classes **2** to **6**. Class **1** is very heterogeneous, and some of its lesions may have been formed by multiple lesions from classes **2** to **6** merging together.

Importantly, the lesion classes can be categorized into two types based on diagnostic implications: isointense lesions (class **2)** and hypointense lesions (class **6**) are *remyelinating* lesions, while PRL lesions (class **3**) and hyperintense lesions (class **5**) are *demyelinating* lesions. In remyelinating lesions, a repair process called "remyelination" helps create new myelin sheaths on demyelinated neurons. Thus, remyelinating lesions indicate a more positive diagnosis for the patient. Demyelinating lesions are more destructive, indicating a worse diagnosis for the patient. Unclassified lesions (class **1**) and hypointense rims are excluded from this categorization. The name, type, and the number of lesions for each class are shown in **Table 1**.

Class	Description	Type	Class size
1	unclassified	/	3664
2	isointense	remyelinating	460
3	PRL	demyelinating	214
4	hypointense rim	/	19
5	hyperintense	demyelinating	841
6	hypointense	remyelinating	71

Table 1: Lesion classes: description, type, and size of class.

3.2 Problem Specification

Being able to differentiate between MRI images of remyelinating lesions and those of demyelinating lesions would provide medical doctors valuable information for MS diagnosis, due to the two types' opposing implications for the state of a patient. A software model that could automatically classify lesions into one of the two types would help doctors by reducing the time required needed for diagnostics, and by potentially improving diagnosis accuracy. We aim to develop a convolutional neural network model that, given a 35x35x35 image of an MS lesion consisting of the FLAIR, QSM, and mask channels, classifies the lesion as being either remyelinating or demyelinating. The input lesions is assumed to belong to classes **2**, **3**, **5**, or **6**, as class **1** and class **4** lesions are considered neither remyelinating nor demyelinating. We present the different parts of our model in the following section.

4. METHODS

We train our convolutional neural network on one part of the dataset – the *training set* – and then test the network on the other part of the dataset – the *test set*. We let the training set consist of 70% of the images and the test set consist of the other 30% (meaning the two sets are disjoint). More precisely, the training set consists of 70% of the images from classes **2**, **3**, **5**, and **6**, while the test set consists of the remaining images from these classes.

4.1 Dataset manipulation

There are two problems imposed by our dataset, especially regarding network training: imbalanced data and the dataset size itself. The imbalance of data is due to there being approximately two times as many demyelinating lesion images (1055) as remyelinating lesion images (531). As our model classifies MS lesions into these types, it would be optimal if the two types had the same number of images in the training set. Otherwise, if the training set is imbalanced, the network might learn to prioritize the lesions of the more numerous type, i.e., the network may incorporate a bias towards demyelinating lesions. This type of bias is undesirable, even if the test set is imbalanced the same way as the training set. This is because such a bias would be a result of dataset statistics, rather than the information and features contained in the images themselves. We solve this problem by simple *oversampling*: we double the number of remyelinating lesion images in the training set by copying each such image. The copies do not introduce new image-specific information into the training set, rather, they prevent the network from learning a bias from data imbalance. We perform oversampling only on the training set and leave the test set imbalanced; this is due to three reasons. First, the test set does not influence the network during training, so it will not introduce a bias. Second, the imbalance is a property of the original dataset, so evaluating the model using the imbalanced test set is remaining faithful to original data. Third, when testing, we are able to see the performance of the network on the remyelinating and demyelinating types separately, in the form of *true positive* and *true negative* rates for each type.

The second dataset problem concerns its size and requires a more complicated solution. Out of the 3664 lesions in total, we consider only 1586 to be remyelinating or demyelinating. This is a small number of images, especially compared to commonly used 2D image datasets such as MNIST (Deng, 2012), CIFAR-10, and ImageNet (Deng et al., 2009), which have tens of thousands or even tens of millions images[2]. In general, medical image datasets contain small numbers

[2] We note that most classification problems defined on those datasets include tens of classes. Binary classification – such as done by our network – is a less demanding task.

of samples (Litjens et al., 2017); our dataset is large in that regard, but its size is still a problem for neural networks due to multiple reasons. First, a larger the training set means more information is available for the network to learn. Second, a network that has the right structure and a large enough number of parameters may be powerful enough to memorize all of the training images, instead of learning the important image features they share. In this case, the network would perform well on the training set but poorly on the test set – a phenomenon called *overfitting*. A larger training set would be harder to memorize. For these reasons, we increase the dataset significantly by using *data augmentation* on the training set.

Data augmentation means increasing the dataset by adding slightly modified copies of existing data, thus producing new data expected to be faithful to the original dataset. In our case, we augment the existing images by using rotation and flip transformations. Because CNNs are not equivariant to rotations (Kim et al., 2020) and flips (Dudar et al., 2018) of images, the augmented images provide the network with more information. We rotate the images by multiples of 90 degrees, i.e., 90, 180, and 270 degrees, around any combination of the three axes, including consecutively using different axes. This choice is due to the fact that rotating by angles that are not multiples of 90 degrees would require interpolation when computing new voxel values, potentially changing the fine-grained information contained in the images. There are 24 rotational symmetries of a cube, so augmenting by all possible rotation combinations produces 24 unique images for every image in the original training set. After rotation, we further augment the training set by flipping the images vertically. A single flip is never equivalent to a series of rotations, so flipping doubles the number of unique images in the training set. On the other hand, a single flip of an image across a different axis (e.g. horizontally) may be achieved by a series of rotations followed by vertical flipping, thus not necessarily producing new unique images. We therefore only perform vertical flipping, to increase the size of the balanced training set $24 \times 2 = 48$ times in total.

4.2 Network architecture

Once oversampling and data augmentation have been performed on the training set, we train the convolutional neural network. We present the architecture that has given the best results in our experiments in **Table 2**. The network consists of the image input layer, three occurrences of a convolutional layer followed by a batch normalization layer, a ReLU activation function, and a max-pooling layer, a fully connected layer of width 2, and the output layers (softmax and cross-entropy layers). The image input layers takes as input images of size $35 \times 35 \times 35 \times c$, where $c \in \{1, 2, 3\}$ is the number of channels used. The first two convolutional layers consist of 64 feature maps, the third convolutional layer consists of 128 feature maps. Increasing the number of feature maps as we move to the deeper convolutional layers is a common practice – perhaps due to the assumption that there are more high-level features than low-level ones. ReLU activation functions (Glorot et al., 2011) have been proven successful in CNNs (Xu et al., 2015).

Layer	Type	Description
1	3D Image Input	$35 \times 35 \times 35 \times c$ images, 'zerocenter' normalization
2	Convolution	64 $3 \times 3 \times 3$ convolutions, stride 1
3	Batch Normalization	
4	ReLU	
5	3-D Max Pooling	$2 \times 2 \times 2$ max pooling, stride 2
6	Convolution	64 $3 \times 3 \times 3$ convolutions, stride 1
7	Batch Normalization	
8	ReLU	
9	3-D Max Pooling	$2 \times 2 \times 2$ max pooling, stride 2
10	Convolution	128 $3 \times 3 \times 3$ convolutions, stride 1
11	Batch Normalization	
12	ReLU	
13	3-D Max Pooling	$2 \times 2 \times 2$ max pooling, stride 2
14	Fully Connected	Size 2
15	Softmax	
16	Classification Output	Cross Entropy

Table 2: Architecture of our convolutional neural network.

4.3 Learning parameters

The parameters used for training the network are as follows. We use Adam (Kingma et al., 2014), a popular stochastic first-order gradient-based optimization algorithm, as our optimization algorithm. Empirical results have shown that Adam has advantages compared to certain other popular optimization algorithms, due to its computational efficiency and numerical robustness (Kingma et al., 2014). We train the network for 15 epochs, using a mini-batch size of 128 images, and shuffling the training set before every epoch. The initial learning rate is 0.01, then decreasing to 0.00316 at epoch 6, and decreasing again to 0.001 at epoch 10. Starting with a larger learning rate allows the network to make large modifications to its parameters, thus being able to change quickly at the beginning of training. In the later epochs, the network has coarsely learned some important image features, and a lower learning rate allows it to then fine-tune those features. We note that all of the training has been done on GPUs, which is a

standard way of accelerating the computationally very costly task of neural network training.

The network has been trained on all possible combinations of the three channels.

5. RESULTS

We present the results of training our network using different combinations of image channels in **Table 3**. Because oversampling was not used in the test set, the set is imbalanced and contains 160 remyelinating and 318 demyelinating lesions. Thus, the baseline accuracy for our test set is 66.66%, as a network that classifies all lesions as demyelinating trivially achieves that accuracy.

The best test accuracies obtained by our network are promising, being around 89% percent; they are highlighted in bold in **Table 3**. The true positive rates are similar (i.e., both larger than 86%) for both remyleniating and demyelinating types in such cases. We notice the following from the table. The accuracies are high if and only if the QSM MRI channel is used, consistently being approximately ∼89%, irrespective of what other channels were used alongside QSM. When the FLAIR MRI and mask channels are used without QSM, the test accuracies are much lower – not surpassing 72%, which is not a significant improvement to the baseline. We conclude that QSM is necessary for the network to successfully classify our dataset, and that the FLAIR and mask channels do not contribute to a better accuracy when QSM is used. This concurs with how the dataset was originally manually classified, as it was QSM that allowed differentiating between the different lesion classes (Rahmanzadeh et al., 2022).

Channels used	Test accuracy
None (*baseline*)	66.66%
Flair	70.09%
QSM	**89.31%**
Mask	69.22%
Flair, QSM	**88.59%**
Flair, Mask	71.01%
QSM, Mask	**89.45%**
Flair, QSM, Mask	**89.00%**

Table 3: Results of training our network using different combinations of channels.

While the FLAIR and mask channels do not contribute to the convolutional neural network itself, the two channels were important in obtaining the dataset. Each image in the dataset was centered around a lesion, and the centering was performed by using the mask of the original MRI images, which indicates the location of a lesion. The mask itself was obtained by using a segmentation neural network on the FLAIR and MP2RAGE MRI images. Therefore, the FLAIR and mask channels were necessary in setting up the model.

6. CONCLUSION

We developed a convolutional neural network model for classifying multi-modal volumetric MRI images of multiple sclerosis lesions into remyelinating or demyelinating types. Due to the imbalance and small size of our dataset for training purposes, we performed oversampling and data augmentation, thus balancing and greatly increasing the dataset. We have compared the results of training the network using different combinations of three image channels. The network achieves a test accuracy of ∼89% whenever the QSM MRI channel is used. The Flair and mask channels do not contribute to a higher accuracy when used alongside the QSM channel, and do not significantly improve the baseline network accuracy when used without the QSM channel.

Traditional computer vision and image processing methods that do not use neural networks seldom achieve significant results in image classification problems. Neural network approaches have dramatically increased the state of the art accuracies on many such problems. However – mostly due to the scarcity of data in medical imaging – such approaches at times still do not achieve adequate results on medical imaging datasets. By achieving a test accuracy of 89%, we have shown that a convolutional neural network approach is a very promising one for our problem. A further improvement to the accuracy would be beneficial to developing trust in the model and to its applicability in aiding medical doctors. Such a software solution may help doctors by automatically classifying images, thus reducing the time and energy needed for diagnosis. We note that our attempts of using more augmentation and larger networks did not result in increasing the presented accuracies, thus implying that a larger dataset might be needed for that purpose.

A large obstacle to deploying a model such as ours in clinical practice would be the difficulty of discerning between images that belong to class **1** of our dataset – images medical doctors were unable to classify due to lesion heterogeneity or presence of artifacts – from those belonging to classes **2** to **6**. Our model assumes that input images belong to classes **2**, **3**, **5**, or **6**; however, most images in our dataset were unclassified ones. Our attempts at constructing a neural network model similar to the one described in Section **5**, with the goal of successfully discerning classified images from unclassified ones, produced no positive results whatsoever. Future work of addressing ways to discard lesions unable to be classified would be necessary for applying our model in clinical practice, especially due to the frequency of such lesion images. Including class **4** lesions in a classification model would also be beneficial, even if such lesions appear infrequently and do not carry large implications for a patient's state.

Finally, a model being able of classifying lesions into each of the **5** classified types, rather than just the remyelinating and demyelinating types, would provide medical doctors with even more information. We conclude that the results of our work are promising, yet further work is needed for applying the model in practice, especially regarding identifying unclassified lesions.

REFERENCES

Al-Saffar, A., Tao, H., & Talab, M.A. (2017). Review of deep convolution neural network in image classification. 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), 26-31.

Ciresan, D.C., Meier, U., Masci, J., Gambardella, L.M., & Schmidhuber, J. (2011). Flexible, High Performance Convolutional Neural Networks for Image Classification. IJCAI.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248-255.

Deng, L. (2012). The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. IEEE Signal Processing Magazine, 29, 141-142.

Dudar, V., & Semenov, V. (2018). Use of symmetric kernels for convolutional neural networks. ArXiv, abs/1805.09421.

Fukushima, K. (2004). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36, 193-202.

Gao, J., Yang, Y., Lin, P., & Park, D.S. (2018). Computer Vision in Healthcare Applications. Journal of Healthcare Engineering, 2018.

Kim, J., Jung, W., Kim, H., & Lee, J. (2020). CyCNN: A Rotation Invariant CNN using Polar Mapping and Cylindrical Convolution Layers. ArXiv, abs/2007.10588.

Kingma, D.P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980.

La Rosa, F., Abdulkadir, A., Fartaria, M.J., Rahmanzadeh, R., Lu, P., Galbusera, R., Barakovic, M., Thiran, J., Granziera, C., & Cuadra, M.B. (2020). Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE. NeuroImage : Clinical, 27.

Litjens, G.J., Kooi, T., Bejnordi, B.E., Setio, A.A., Ciompi, F., Ghafoorian, M., Laak, J.V., Ginneken, B.V., & Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

Lundervold, A.S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. Zeitschrift fur medizinische Physik, 29 2, 102-127.

Minaee, S., Boykov, Y., Porikli, F.M., Plaza, A.J., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image Segmentation Using Deep Learning: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44, 3523-3542.

Rahmanzadeh, R. (2022). Quantitative susceptibility mapping identifies remyelinated, chronic active and chronic inactive multiple sclerosis lesions. Preprint.

Reich, D.S., Lucchinetti, C.F., Calabresi, P.A., Jan. (2018). Multiple sclerosis. N. Engl. J. Med. 378 (2), 169–180.

Yanase, J., & Triantaphyllou, E. (2019). A systematic survey of computer-aided diagnosis in medicine: Past and present developments. Expert Syst. Appl., 138.

Zhao, Z., Zheng, P., Xu, S., & Wu, X. (2019). Object Detection With Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems, 30, 3212-3232.