Technische Universität München
Fakultät für Medizin

A lesion-level deep learning approach to predicting enhancing lesions from non-enhanced images in Multiple Sclerosis patients

Nikhil Sasidharan

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung eines

Doktors der Medizin (Dr. med)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Florian Eyer

Prüfer*innen der Dissertation:

1. Priv. Doz. Dr. Benedikt Wiestler
2. Priv. Doz. Dr. Matthias Elstner

Die Dissertation wurde am 31.01.2023 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 18.07.2023 angenommen.

Table of Contents

List of Figures

List of Tables

Summary

Multiple Sclerosis is an auto-immune disorder of the central nervous system characterized by inflammatory demyelination and neurodegeneration of the brain. Neuroradiologic assessment of MR images plays an important role in the diagnosis of the disease as well as in the long-term monitoring of disease activity. Gadolinium-based contrast agents are often used to identify inflammatory active lesions and serve as a proxy for active disease. Such contrast agents have been shown to accumulate in the body leading to potential toxic effects. Nevertheless, since MS patients are typically young at disease onset and require regular follow-up scans, our goal was to develop a lesion-level deep learning approach to predict lesion enhancement in non-enhanced MRI scans of MS patients. Such an approach would ideally negate the requirement to use contrast agents. In total, 485 consecutive patients with 1034 exams from a prospective observational cohort of Multiple Sclerosis patients were included in this analysis (scanned between 01/2008 and 09/2017). 10,311 lesion-level patches were generated from 297 patients (Mean age: 34.8 years ± 9.7, Ratio Female: Male 196:101) to train a deep convolutional neural network (85% training, 15% validation) to discriminate between enhancing and non-enhancing lesions. Manual annotation of individual lesions enhancement on T1-weighted images after contrast administration served as the ground truth. This network was then tested on the remaining 188 patients (Mean age: 33.9 years ± 9.1, Ratio Female: Male 130:58). Network performance was statistically assessed by AUC, sensitivity, specificity and accuracy analysis. On the lesion-level, an AUC of 0.905 was observed. A pre-specified probability threshold of 0.75 allowed for identification of enhancing lesions with a sensitivity of 47.65%, specificity of 97.27% and accuracy of 95.27%. On patient-level, the network exhibited a sensitivity of 82.76%, specificity of 66.15% and accuracy of 71.28%. An external data set of 29 patients was additionally used to validate and confirm these results. Our study revealed that a deep learning approach at the lesion-level is effective in predicting contrast-enhancing lesions from non-enhanced MRI images of patients with MS.

Zusammenfassung

Multiple Sklerose ist eine Autoimmunerkrankung des zentralen Nervensystems, die durch entzündliche Demyelinisierung und Neurodegeneration des Gehirns gekennzeichnet ist. Die neuroradiologische Beurteilung von MR-Bildern spielt eine wichtige Rolle bei der Diagnose der Krankheit sowie bei der langfristigen Überwachung der Krankheitsaktivität. Kontrastmittel auf Gadolinium Basis werden häufig zur Identifizierung entzündlich-aktiver Läsionen verwendet und dienen als Indikator für eine aktive Erkrankung. Es hat sich gezeigt, dass sich solche Kontrastmittel im Körper anreichern und zu potenziell toxischen Wirkungen führen können. Da MS-Patienten bei Ausbruch der Krankheit in der Regel jung sind und regelmäßige Nachuntersuchungen benötigen, war es unser Ziel, einen Deep-Learning-Algorithmus auf Läsionsebene zu entwickeln, um die entzündlich-aktiven Läsionen in nativen MRT-Scans von MS-Patienten vorherzusagen. Ein solcher Ansatz würde im Idealfall den Einsatz von Kontrastmitteln überflüssig machen. Insgesamt wurden 485 konsekutive Patienten mit 1034 Untersuchungen aus einer prospektiven Kohorte von Multiple-Sklerose-Patienten in diese Analyse einbezogen (gescannt zwischen 01/2008 und 09/2017). Von 297 Patienten (Durchschnittsalter: 34,8 Jahre ± 9,7, Verhältnis weiblich:männlich 196:101) wurden 10.311 Patches auf Läsionsebene generiert, um ein neuronales Netzwerk zu trainieren (85 % Training, 15 % Validierung), um zwischen potenziell Kontrastmittel anreichernden und nicht-anreichernden Läsionen zu unterscheiden. Die manuelle Annotation der Läsionen auf Basis der T1-gewichteten Bildern nach Kontrastmittelgabe diente dabei als Grundwahrheit. Dieses Netzwerk wurde dann an den verbleibenden 188 Patienten (Durchschnittsalter: 33,9 Jahre ± 9,1, Verhältnis weiblich:männlich 130:58) getestet. Die Leistung des Netzwerks wurde anhand von AUC-, Sensitivitäts-, Spezifitäts- und Genauigkeitsanalysen statistisch bewertet. Auf der Ebene der Läsionen wurde dabei eine AUC von 0,905 beobachtet. Bei einem vorgegebenen Wahrscheinlichkeitsschwellenwert von 0,75 konnten anreichernde Läsionen mit einer Sensitivität von 47,65 %, einer Spezifität von 97,27 % und einer Genauigkeit von 95,27 % identifiziert werden. Auf Patientenebene wies das Netzwerk eine Sensitivität von 82,76 %, eine Spezifität von 66,15 % und eine Genauigkeit von 71,28 % auf. Zusätzlich wurde ein externer Datensatz von 29 Patienten zur Validierung verwendet, der diese Ergebnisse bestätigte. Unsere Studie zeigte, dass ein Deep-Learning-Ansatz auf Läsionsebene bei der Vorhersage von kontrastanreichernden Läsionen aus nativen MRT-Bildern von Patienten mit MS ein vielversprechender Ansatz ist.

Abbreviations

MS – Multiple Sclerosis
GBCA – Gadolinium Based Contrast Agents
CIS – Clinically Isolated Syndrome
RRMS – Relapsing Remitting MS
CNN – Convolutional Neural Network
DL – Deep Learning
ML – Machine Learning
DMT – Disease Modifying Therapy
FLAIR – Fluid-attenuated inversion-recovery
LST – Lesion Segmentation Tool
ANT – Advanced Normalization Tool
SPMS – Secondary progressive MS
ROC – Receiver Operating Characteristic
AUC – Area Under the Curve
TP – True Positive
TN – True Negative
FN – False Positive
CE- – Manually labelled as non-enhancing
CE+ – Manually labelled as contrast-enhancing
Pred CE- – Predicted to be non-enhancing
Pred CE+ – Predicted to be enhancing
EDSS – Expanded Disability Status Scale

Introduction

Multiple Sclerosis (MS) is an auto-immune disease of the central nervous system characterized by inflammatory demyelination followed by axonal and neuronal degeneration (McGinley et al., 2021; Reich et al., 2018). More than 2.5 million people are affected worldwide (Reich et al., 2018; Tullman, 2013). MS patients often initially present with a neurological episode that lasts at least 24 hours, which is referred to as Clinically Isolated Syndrome (CIS) (Tafti et al., 2021). As disease progresses, MS patients can be divided into 4 groups based on their disease course: relapsing-remitting MS (RRMS), primary-progressive MS (PPMS), secondary-progressive MS (SPMS) and in 5% of cases as Progressive-relapsing MS (PRMS) (Figure 1) (Lublin & Reingold, 1996; Tafti et al., 2021). RRMS is the most prevalent form and is defined by a sequence of attacks followed by recovery over time, PPMS is characterized by a gradual progression of symptoms after initial onset, SPMS patients present with a gradual deterioration after presentation with initial RRMS and PRMS patients have gradual decline in parallel to recurrent relapses (Figure 1) (Tafti et al., 2021).

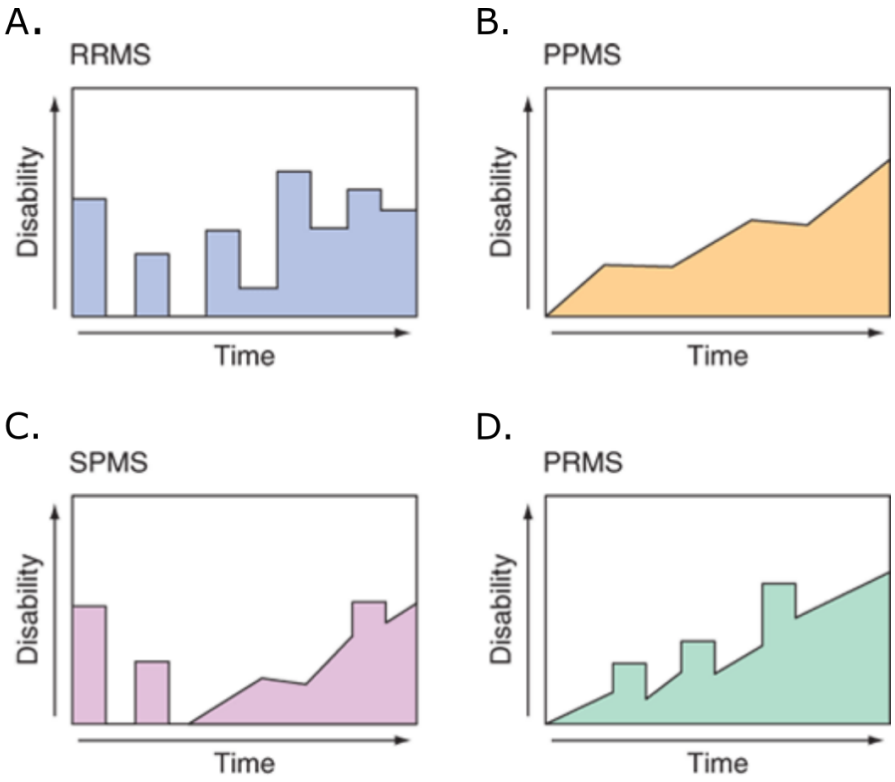


Figure 1. MS Patients can be divided in 4 groups based on the course of the illlness.

A. Relapsing-Remitting MS (RRMS) B. Primary-Progressive MS (PPMS) C. Secondary-Progressive MS (SPMS) D. Progressive-Relapsing MS (PRMS). Image modified from (Stephen et al., 2016).

Although the exact cause of MS remains unknown, from a pathology perspective, MS lesions in the brain can be characterized as being pre-active, active or inactive (Jonkman et al., 2015; Reich et al., 2018; Van Der Valk & De Groot, 2000). Pre-active lesions display aggregations of microglia, but display no changes in myelination, active lesions display partial demyelination with the presence of immune cells and inactive lesions show complete demyelination as well as an absence of immune cells (Van Der Valk & De Groot, 2000). At a given point in time, a patient

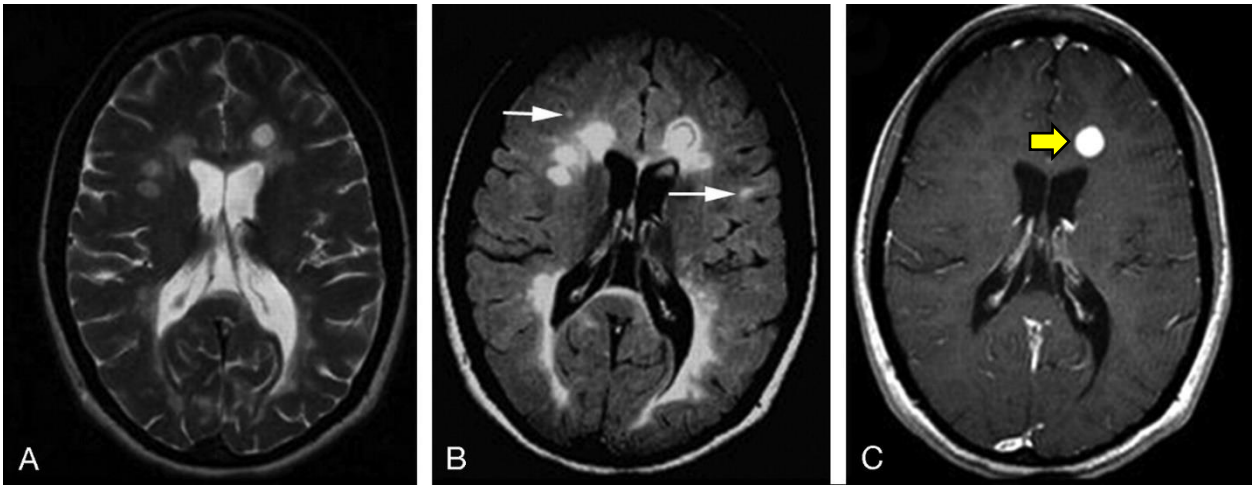


Figure 2. MR scans of a patient with active RRMS.

A. T2w image B. FLAIR image C. Contrast-enhanced T1w image. Both enhancing and non-enhancing lesions are observed depicting the various stages of lesion progression. Small lesions are only visible on FLAIR and depicted with white arrows. Contrast enhancing lesions are shown with yellow arrows. *Image modified from (Ge, 2006).*

may have a combination of several lesion types with a weak correlation with clinical symptoms (Jonkman et al., 2015; Li et al., 2006). Neuroradiographic assessment plays an increasingly important diagnostic role in the diagnosis of lesions *in vivo* as described in the original McDonald criteria (Mcdonald et al., 2001). Diagnostic criteria have been refined such that an MS diagnosis can be made solely through a single contrast-enhanced study, which reveals dissemination in time through the simultaneous presence of enhancing and non-enhancing lesions as well as dissemination in space by identifying lesions in several pre-defined anatomical areas (Thompson et al., 2018).

Active inflammatory lesions are visualized through intravenous injections of Gadolinium-based contrast agents (GBCA) coupled to a chelating agent (Saade et al., 2018). Contrast-enhancement is indicative of active inflammatory processes in and surrounding a lesion with a

disruption of the blood-brain-barrier (BBB) (Brück et al., 1997; Saade et al., 2018). Figure 2 is a sample MRI sequence (T2w, FLAIR and contrast-enhanced T1w images) of an RRMS patient with an active contrast enhancing lesion as well as non-enhancing lesions (Ge, 2006). The identification of enhancing lesions has been shown to be more sensitive for determining new disease activity than clinical monitoring (Barkhof et al., 1992). These lesions, greatly varying in shape and size, appear in the form of homogenous enhancing nodules and over time develop into enhanced ringlike structures (Ge, 2006). Due to the chronic nature of MS, patients are often monitored through repeated contrast-enhanced MR imaging over many years.

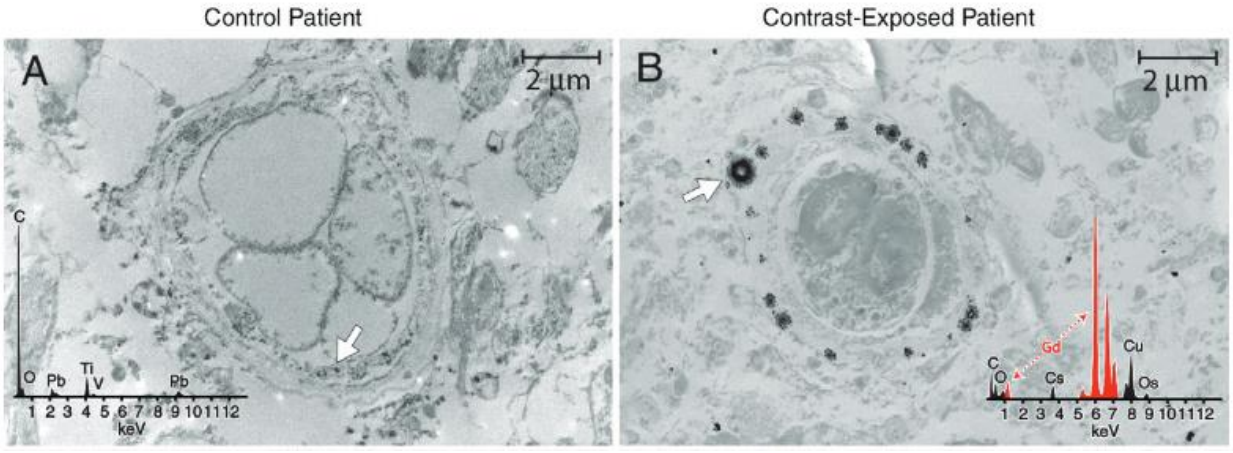


Figure 3. Electron microscopy of post-mortem tissues show accumulation of GBCAs.

Dentate nuclei from A, control patient and, B, contrast group patient. X-ray spectra of dense foci (arrow); Peaks in red indicate gadolinium. C= carbon, Cs = cesium, Cu = copper, Gd = gadolinium, O = oxygen, Os = osmium, Pb = lead, Ti = titanium, V = vanadium. *Image modified from (Mcdonald et al., 2015).*

The use of GBCAs have been shown to be associated with toxic health effects (Rogosnitzky & Branch, 2016). Repeated exposure to Gadolinium can cause adverse effects at a range of 0.03% to 2.4% with detrimental effects documented in the central nervous system as well as in other parts of the body as seen through nephrogenic systemic fibrosis (NSF) (Kanal et al., 2014; Kanda et al., 2016; Rogosnitzky & Branch, 2016). Development of NSF can lead to severe disablement and death (Swaminathan et al., 2008). The continuous use of contrast agents have been shown to be detrimental to the body. Furthermore, GBCA has shown to accumulate in the brain also in patients with normal kidney function (Errante et al., 2014; Kanda et al., 2014).

Electron microscopic and mass spectrometric analysis of post-mortem brain tissue has localized intracranial Gadolinium deposition to capillary endothelium as well as neural interstitium of the dentate nucleus and globus pallidus (Figure 3) (Mcdonald et al., 2015). Kanda et al. further

demonstrated that the degree of hyperintensity correlated with the number of previous Gadolinium administrations (Kanda et al., 2014). Therefore, various organizations such as the International Society for Magnetic Resonance in Medicine and the Consortium of MS Centers have advocated caution in administration of GBCA (Gulani et al., 2017). This has led to the development of alternative efforts to avoid application of GBCAs. Nonetheless, identification of active inflammation through contrast-enhancement still carries clinically relevant information.

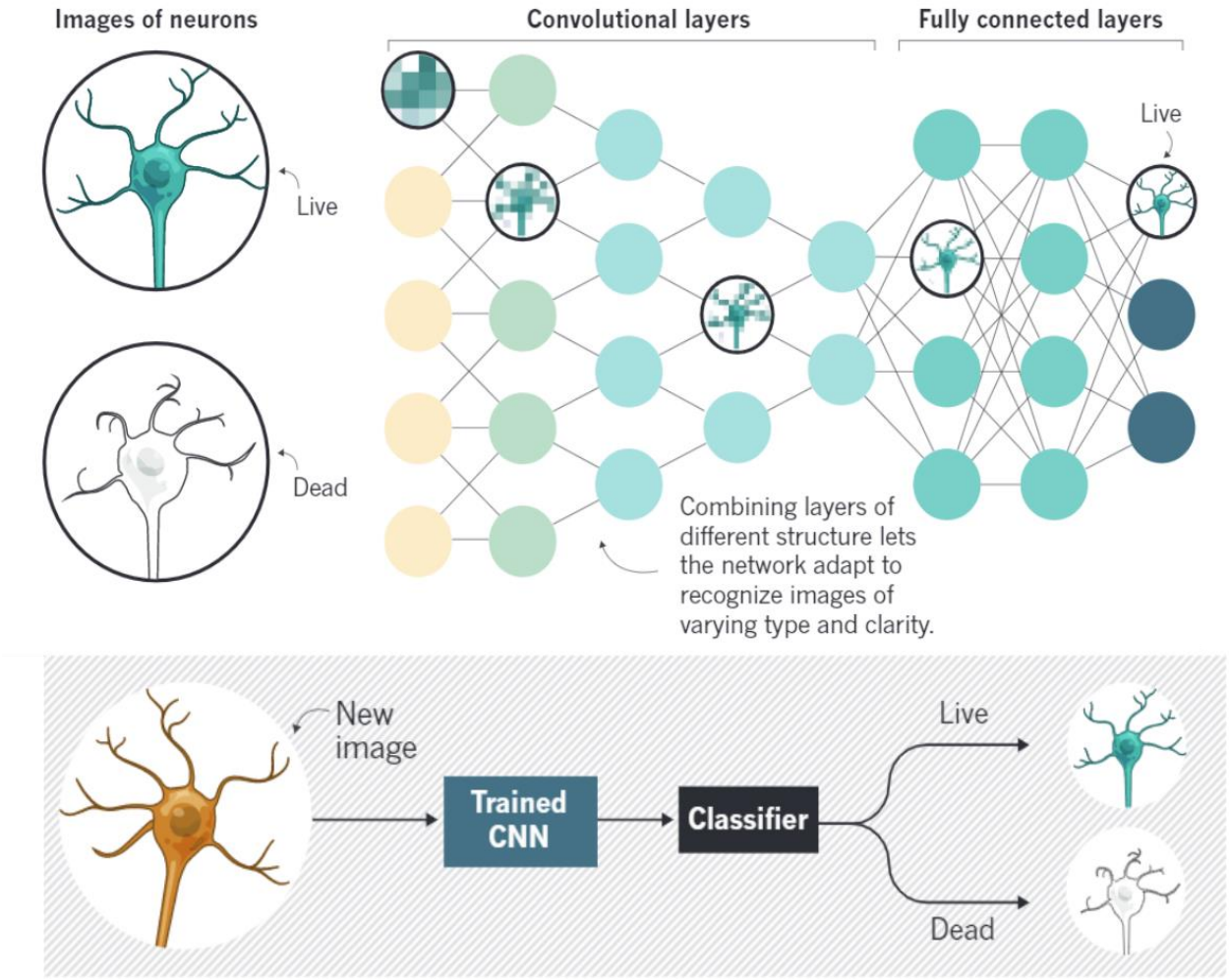


Figure 4. Overview of a Deep Learning algorithm.

DL algorithms consist of two major steps: A training step and a testing step. In the first step, a set of images are used as input to generate a network (Convolutional Neural Network, CNN) images are identified and clustered based on specific properties of the images. Then in a second step, a new image is run through the CNN to classified based on the properties attributed in the network. *Image modified from (Webb, 2018).*

Over the last several years, deep learning (DL) algorithms have proven to be effective in solving complex problems in the fields of biology and medicine. Modern DL algorithms in image analysis are based on convolutional neural networks (CNNs) generated to identify hidden patterns through training on large datasets (Webb, 2018). DL networks are characterized by the presence of multiple layers between input and output (Huang et al., 2017), allowing them to learn complex nonlinear mappings between input and output variables. After a network has been trained a new dataset can be applied to it in order to extract meaningful information (Figure 4). DL approaches have been used in the mining of genomic data to identify Single Nucleotide Polymorphisms, for

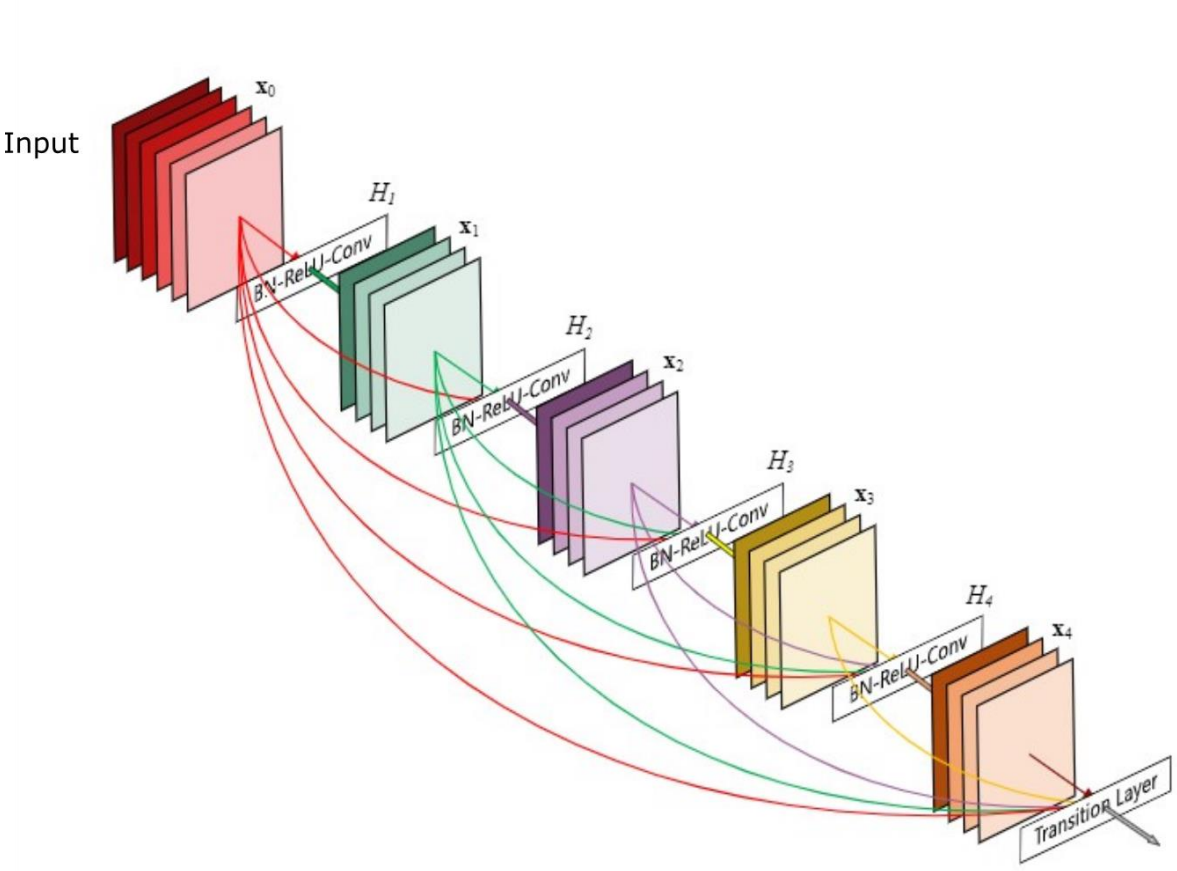


Figure 5. Schematic representation of an interconnected DenseNet Block.

The block consists of 5 layers that receive input from all other layers. *Image modified from (Huang et al., 2017).*

virtual drug screening and for the stratification of disease types (Carpenter et al., 2018; Poplin et al., 2018; Webb, 2018)**.** Recently DL approaches have also been used in the field of radiology for lesion detection, segmentation (Li et al., 2018) and classification (Narayana et al., 2020) in MS.

We hypothesized that contrast-enhancing, “active” MS lesions show a distinct imaging phenotype on non-enhanced images, which a neural network can recognize.

To test our hypothesis, we trained and validated a deep learning classifier on a segmented lesion-level library of MS patients with contrast-enhancing and non-enhancing lesions to learn the characteristics of a typical contrast-enhancing lesion. To generate the classifier we made use of a DenseNet, a type of network architecture where all layers are interconnected in a feed-forward structure as seen in figure 5 (Huang et al., 2017). DenseNets have been proposed to significantly improve deep learning networks due to their simplified connectivity pattern, reduced redundancy and improved feature propagation (Huang et al., 2017).

Our second objective was to apply the DenseNet to a test data set to predict lesions that would be contrast-enhancing on follow-up and assess accuracy of the network. In a further step, we sought to determine if we could predict which patients would have enhancing lesions in a follow up scan using only the baseline images. We accomplished this using the results of our classifier in an exploratory analysis to generate a multivariate model. The results of the model were then compared with the follow up scans to determine if the predictions were significantly associated with the manual annotations.

Materials and Methods

Study participants

For this study a secondary analysis of a monocentric prospective observational cohort was conducted. Data was collected from MS patients treated in the Neurology Department of the Klinikum rechts der Isar. Our local institutional review board reviewed and approved this study. Additionally participants were informed about the study and provided written consent to participate.

Our inclusion criteria considered all patients newly-diagnosed with Clinically Isolated Syndrome (CIS) or Relapsing Remitting MS (RRMS) that have been scanned at a Philips Achieva 3T scanner with a standardized scan protocol. The patient collective was restricted to a time period between 01/2008 and 09/2017 and only those with a follow-up scan at least within a 32 months period from their first image were considered. Patients treated with Disease Modifying Therapies (DMT) for a period longer than 6 months at the point of presentation for their first scan or patients diagnosed more than 3 years prior to or more than 1 month after their first scan were excluded from the study. Furthermore, we also excluded all exams without Gadolinium contrast-enhanced T1-weighted sequences (n=160), resulting in a final total of 485 patients with 1034 exams. For training and validation our deep learning algorithm, we used patients with at least four exams (n = 297 participants, 846 exams; 85% training, 15% validation). Additionally, lesions were included from both baseline and follow-up studies for these patients. Subsequent to training and validation, in order to test the algorithm we used the remaining patients in the cohort and trained the algorithm using only baseline images of lesions. (n = 188 participants).

MR image acquisition, processing and annotation

For this project, a 3.0-T system (Achieva; Philips Healthcare, Best, the Netherlands, Software version 5.1, 8-channel coil) MRI was used for acquiring images. A three-dimensional fluid-attenuated inversion-recovery (FLAIR) sequence as well as three-dimensional T1-weighted sequences performed before and at least 5 minutes after administration of 0.1 mmol/kg gadolinium-based contrast material (Dotarem [gadoterate meglumine]; Guerbet, Villepinte, France) was included during image acquisition. For three-dimensional FLAIR images were acquired using the following parameters: acquired voxel size, 1.03x1.03x1.5 mm^3; repetition time 10 000 msec, echo time 140 msec; acquisition time, 5 minutes; plane, axial. For three-dimensional

T1-weighted gradient-echo imaging, the following parameters were used: acquired voxel size, 1x1x1 mm^3; repetition time 9 msec, echo time 4 msec; acquisition time 6 minutes; plane sagittal.

The baseline exam from each participant was segmented using the lesion segmentation tool (LST) (Schmidt et al., 2012) and in the follow-up exams new or enlarging lesions were detected using the longitudinal version of LST (Schmidt et al., 2019). The 2017 McDonald criteria was used to automatically categorize the lesion locations anatomically (Thompson et al., 2018) in periventricular, (juxta)cortical, subcortical and infratentorial locations, using a non-linear ANTs-based registration (Avants et al., 2008) to the MNI152 atlas template. In parallel, T1w subtraction maps were generated by subtracting the non-enhanced T1w image from the contrast-enhanced T1w image. In a following step, baseline segmentations were verified and manually corrected (Yushkevich et al., 2006) for accuracy of voxel-level segmentation and localization by a medical student and a board certified neuroradiologist with 3 years of experience. In addition, the presence of contrast-enhancement was annotated for each lesion using the T1w subtraction maps. In a final step, all lesions with a volume less than $7mm^3$ were removed, as these represent with a high likelihood unspecific lesions (Grahl et al., 2019). All completed segmentations were reviewed by a neuroradiologist with 17 years of experience in neurological imaging.

For images of follow-up exams, longitudinal FLAIR subtraction maps were calculated and used as a reference for manual corrections of the LST longitudinal segmentation as well as determination of the localization of new or significantly enlarged lesions as described previously (Eichinger et al., 2019; Eichinger et al., 2017). Similar to the baseline images, we calculated here contrast vs. non-contrast T1w subtraction maps and classified individual lesions as contrast-enhancing or non-enhancing. As with the baseline images, a manual correction of the follow-up exams were performed by several specifically trained residents as well as a board-certified neuroradiologists and corrected by one experienced board-certified neuroradiologist.

Prior to subsequent patch extraction, all HD-BET skull-stripped (Isensee et al., 2019) images and lesions maps were rigidly transferred to the MNI152 space using ANTs in 1x1x1 mm^3 resolution. The images were winsorized, within the brain mask, between the 0.1st and 99.9th percentile. The intensities were then re-scaled to an interval of [0;1]. For each individual lesion, in addition to the total volume (in mm^3), the 75th percentile of signal intensity in the T1w subtraction maps was also noted.

Validation using external cohort

In order to validate our findings, an independent, external data set comprising 30 MS patients was downloaded from the University of Ljubljana (Lesjak et al., 2018). For all of these patients, FLAIR images and T1w images before and after contrast agent administration were available with the exception of one patient. Additionally, a corresponding manual lesion segmentation was also provided. Using T1w subtraction maps, lesion enhancement was assessed by a neuroradiologist with 10 years of experience. Steps for image processing were identical to those performed on our internal data.

Deep learning architecture and training

Huang et al. proposed, in 2017, DenseNets as a means to solve vanishing gradient problems and to train deep convolutional neural networks for the classification of images (Huang et al., 2017). As this approach met our requirements, we decided to use this architecture for our image classification task. In order to optimize information in the foreground and at the same time to prevent the need for fully segmented lesion masks as is commonly used in radiomics models, we decided to utilize three-dimensional patches extracted around the center of mass of each lesion with a voxel size of 16x16x16 as input. For each individual lesion, a concatenation of the FLAIR and the non-enhanced T1w patches was used as input for the network. The network used in our study was composed of three dense blocks including 6, 12 and 18 convolutional layers. In our approach, we start directly with a full resolution dense block which is in contrast to the original implementation where an initial down-convolution of the input is performed. A growth rate of 32 was set, and after the first and second block a transition block with a reduction of 0.5 was carried out. After the third (final) dense block, global average pooling was performed along with a dense layer with a single sigmoid output neuron. Batch normalization and leaky ReLU activation was used for all convolutional layers. $\lambda 2$ weight regularization was applied to all convolutional layers to better control overfitting. Table 1 summarizes the network architecture.

Adam optimizer was used to perform training and binary cross-entropy loss for 250 epochs. Training was conducted with a polynomial learning rate decay from 0.01 to 0.0001. Considering that 90% of lesions were non-enhancing, we adjusted for the class imbalance by systematically oversampling enhancing lesions to ensure that they composed 50% of samples in each batch (batch size 96). Images were augmented through changes to image morphology and

content. This included flipping, rotating, shifting as well as changes in intensity in the form of gaussian smoothing and random noise. A total of 5 training runs were performed in order to ensemble predictions. Three models were saved for each run: These included the model with highest validation AUC, the model with the highest validation F1 score and the final model after 250 epochs. This led us to have a total of 15 models. An Nvidia P6000 GPU was used for training (Nvidia, Santa Clara, California) and Keras/Tensorflow 2.4 for implementation (Abadi et al., 2016).

The code for generating the 16x16x16 input patches and for defining the model architecture and training runs has been deposited in our Github repository can be accessed using the following link: https://github.com/CompImg/ms_ce_prediction.

Table 1. Overview of DenseNet Architecture utilizing 16x16x16x2 input patches.
ks = kernel size, s = strides

Layer	Architecture	Output size
Dense Block 1	[ks=1,s=1 conv, ks=3,s=1 conv]*6	16x16x16
Transition 1	[ks=1, s=2 conv]	8x8x8
Dense Block 2	[ks=1,s=1 conv, ks=3,s=1 conv]*12	8x8x8
Transition 2	[ks=1, s=2 conv]	4x4x4
Dense Block 3	[ks=1,s=1 conv, ks=3,s=1 conv]*18	4x4x4
Classification	Global average pooling, single sigmoid output	1x1

Prediction of Lesion enhancement and statistical analysis

An average of all predictions for each model type (highest validation AUC, highest validation F1 score, final model) from the five runs, yielded three predicted probabilities per input lesion. Due to the class imbalance, we anticipated a potential tendency for false-positive labelling during validation. Consequently, we set a probability threshold of 0.75 for each prediction in order for it to classified as “enhancing”, and additionally conducted majority voting. Using this approach

it was necessary that at least two models predict a lesion to be “enhancing” with a probability greater than 0.75 in order for a final classification of a lesion as “enhancing”.

In a next step, we calculated prediction probabilities for the internal test set and the external data set from the University of Ljubljana. Each individual lesion was classified in a binary manner as described above. Classification performance was evaluated from the prediction probabilities using the AUC of a receiver operating characteristic curve. Subsequently, diagnostic accuracies using confusion matrices and derived measures i.e. sensitivity or specificity were calculated for binary lesion prediction. In order to understand accuracy on a patient level, we aggregated data on a participant-wise basis. A participant was determined to have enhancing lesions if our model predicted one or more lesions to be “enhancing” and/or if the participant truly had enhancing lesion in the ground truth. Additionally, we explored lesion volume data from the segmentation maps and correlated this with lesions predicted to be enhancing or non-enhancing to determine any potential relationships.

In a final step, to determine the value of our predictions for later disease activity, we compared lesions predicted to be enhancing with truly enhancing lesions. For this purpose, we searched for new or significantly enlarged lesions for each participant using subsequent follow-up scans. These data were then subjected as a binary dependent variable to multivariate logistic regression models.

Results

Patient demographics and flowchart

A description of the patients that participated in this study is summarized on Table 2. As mentioned in the materials and methods, patients with Clinically Isolated Syndrome (CIS) or Relapsing Remitting MS (RRMS) were used for the study. As expected, patients in the training and testing data sets are very similar to each other (all comparisons $p > 0.25$), except for a slightly lower EDSS as baseline in the testing cohort ($p = 0.04$). Patients from the external data had a higher EDSS and were on average older. Additionally, a small proportion of patients were diagnosed with secondary progressive MS (SPMS) (3 / 29).

Based on our inclusion criteria, a total of 774 patients were initially identified to fit our study. Among these patients, we excluded 289 due to scans being outside of the time window specified for the study and/or due to missing contrast enhancing images. In the end, a total of 485 patients were used in the final analysis. These patients were diagnosed with either the CIS or RRMS forms of Multiple Sclerosis. A flow diagram of the patients included in the study is summarized n Figure 6.

Table 2. Overview of patient demographics in study.
SD: standard deviation, CIS: clinically isolated syndrome, RRMS: relapsing-remitting MS, SPMS: secondary progressive MS; EDSS: Expanded Disability Status Scale, IQR: inter-quartile range

	Internal training/validation data (n = 297 participants)	**Internal test data (n = 188 participants)**	**External test data (n = 29 participants) (*Lesjak et al., 2018*)**
Age (years, mean ± SD)	34.8 ± 9.7	33.9 ± 9.1	39.3 ± 10.1
Sex (f, m)	196 / 101	130 / 58	22 / 7
MS type (CIS / RRMS / SPMS)	110 / 187 / 0	78 / 110 / 0	3 / 24 / 2
Mean disease duration at baseline (years, mean ± SD)	0.3 ± 0.7	0.4 ± 1.4	Not provided
EDSS at baseline (median ± IQR]	1.5 (1.0 – 2.0)	1.0 (0 – 2.0)	2.0 (1.5 - 4.0)

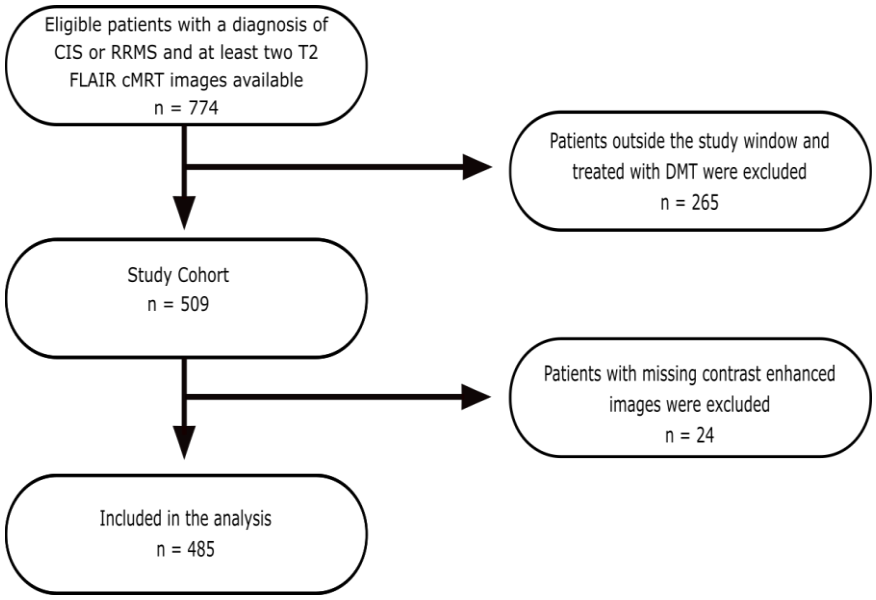


Figure 6. Flow diagram providing overview of study participants.

A total of 774 patients were eligible based on diagnostic and radiologic inclusion criteria. All patients with scans outside the desired timeline as well as patients with missing images were excluded in final analysis. The final cohort consisted of 485 patients.

Manual segmentation and annotation of lesions

Initial LST segmented lesions served as the baseline for the study. Subsequently, lesions were manually annotated to be either enhancing or non-enhancing. As a proof of principle, images manually annotated to be enhancing display a higher T1w subtraction signal compared to those annotated to be non-enhancing (Figure 7), (unpaired t test, $p < 0.001$). Interestingly, lesions that were annotated to be enhancing were also found to be larger in volume when compared to the non-enhancing images (unpaired t test, $p < 0.001$).

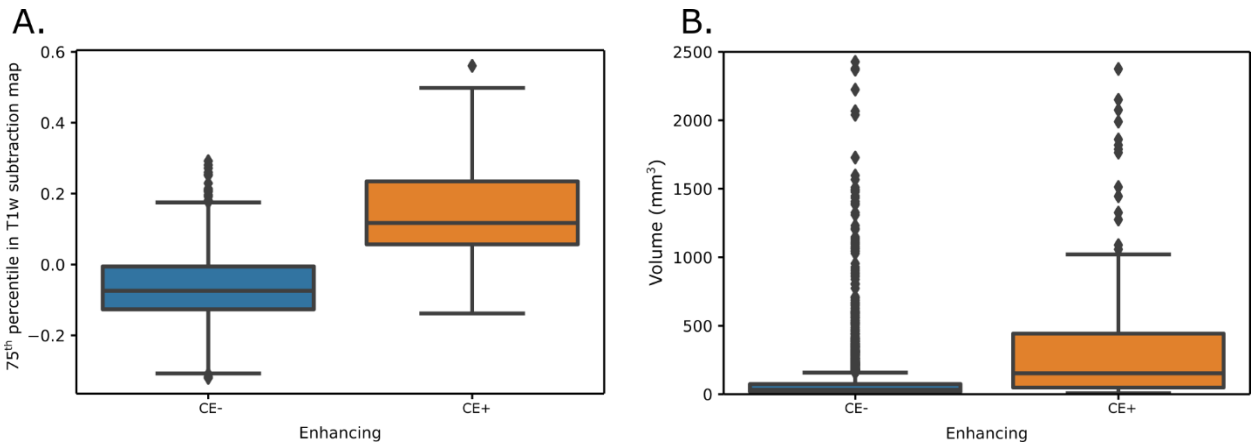


Figure 7. Detecting enhancing lesions through manual annotation.

As expected, results of manual annotation demonstrated that humans can identify and differentiate enhancing lesions from non enhancing ones. A. The top 75% of the manually annotated lesions displayed a higher T1w subtraction signal in the set of lesions identified as contrast-enhancing. B. Volumetric analysis of these lesions show that they are larger in size compared to non-enhancing lesions

Deep learning algorithm to identify enhancing lesions

In order to identify lesions that are potentially enhancing, we used a lesion-level deep learning approach using the manually annotated enhanced images as ground truth images. a total of 10,311 lesions (981 enhancing lesions, 9,330 non-enhancing lesions) from 846 MR exams of 297 participants were used to train the network. Among these images, 85% were used for training and 15% were used for validation of images. To verify that our algorithm was indeed learning from our test images, we generated loss curves for each of the 5 learning runs. As shown in figure 8, each of the loss curves demonstrate a high learning rate.

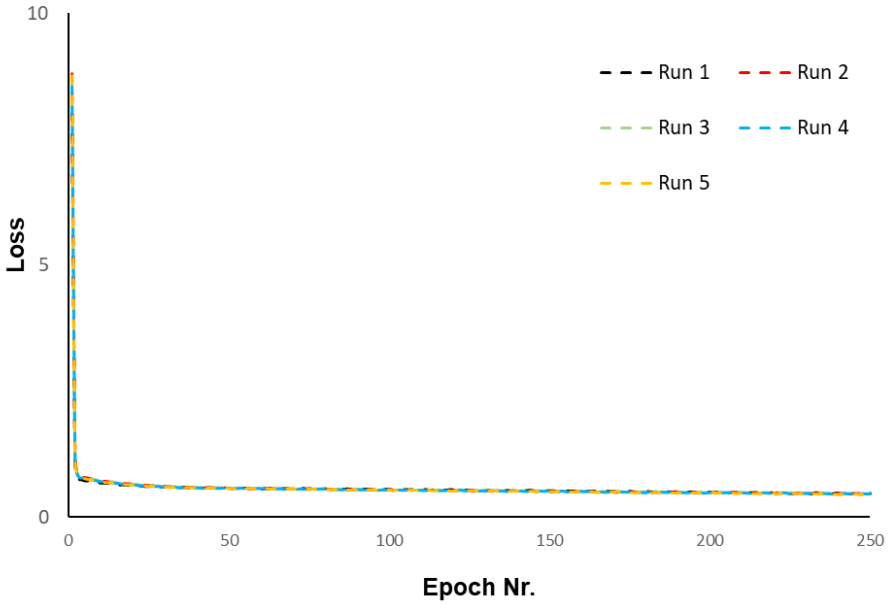


Figure 8. Loss vs. Epoch curve of 5 training runs.

The loss curves demonstrate that the network has a high learning rate. Please note that the curves have been stacked to allow for easier visibility of the individual runs. For graphs of individual runs, refer to appendix.

On application of the resulting algorithm ensemble on our test data set of 188 participants with 4,232 lesions (170 enhancing lesions, 4062 non-enhancing lesions) we obtained an AUC of 0.905 for predicting CE at the lesion level (Figure 9).

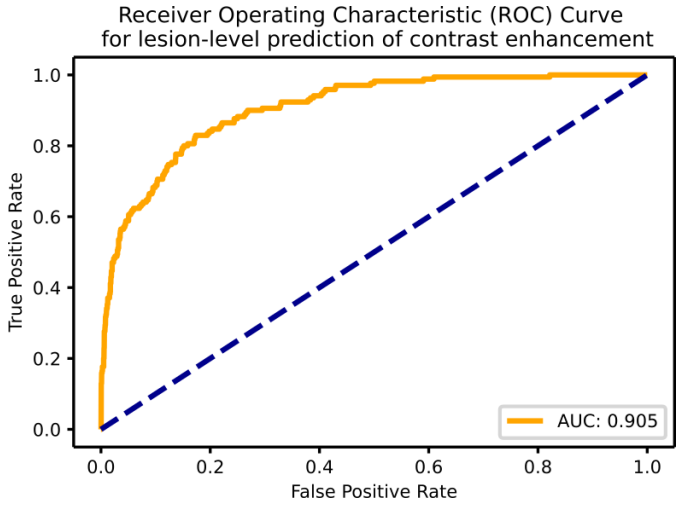


Figure 9. ROC curve.

Utilizing various probability thresholds, ROC show the correlation of the true positive rate vs the false positive rate for detection of enhancing lesions. The dashed blue line indicates potential outcomes based on chance. Orange line shows the average ROC. Mean area under the receiver operating characteristic curve (AUC) is 0.905.

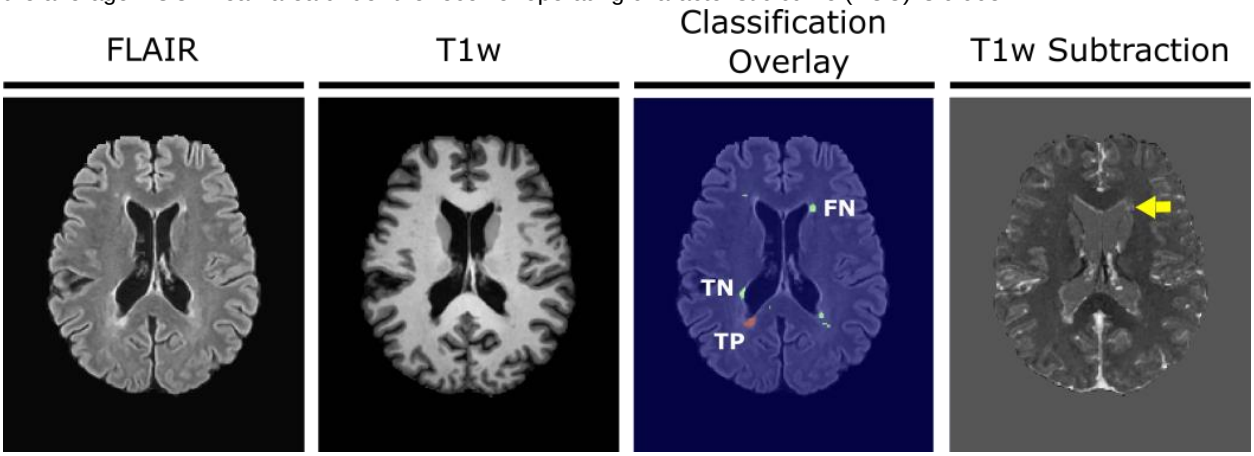


Figure 10. Selected sample of Lesion-wise classification.

All lesions classified by the algorithm were compared to the groundtruth to verify enhancement. Subsequently, lesions were then annotated as either true positive (TP), true negative (TN), false negative (FN) and false positive (not shown). Yellow arrow shows a lesion that was classified to be a FN lesion.

Taking advantage of majority voting on a probability threshold of 0.75, as outlined in the Methods section, we showed a binary accuracy of 95.27% (4032 of 4232 lesions correctly classified) on lesion level. Figure 10 shows a sample visual depiction of how the lesions are classified through our algorithm. On a lesion-level, specificity was determined to be very high (97.27%), while sensitivity was relatively lower (47.65%). On a patient-level, where patients are dichotomized as either having no enhancing lesions (true or predicted) or at least one enhancing

lesion (true or predicted), accuracy was 71.28% (134 of 188 patients correctly classified), sensitivity 82.76% and specificity 66.15%.

In order to better understand enhancing lesions, we looked at additional lesion properties in more detail. It was determined that lesions predicted to be contrast-enhancing indeed displayed significantly higher T1w subtraction map intensities, despite our model never seeing contrast-enhanced T1w images (Figure 11A). This difference was determined to be statistically highly significant (unpaired t test, $p < 0.001$). Furthermore, a difference in lesion volume was also determined to be significant between both groups (unpaired t test, $p < 0.001$), with lesions predicted to be enhancing possessing a greater volume (Figure 11B). Surprisingly, a comparison with the ground truth yielded a similar effect: Enhancing lesions were found to be significantly larger in comparison to non-enhancing lesions (unpaired t test, $p < 0.001$). It is likely that this association is also partly responsible for samples being misclassified. From our observations, samples falsely predicted to be non-enhancing (false-negative) were significantly smaller in comparison to those wrongly predicted to be enhancing (false-positive) (Figure 11C; unpaired t test, $p = 0.004$).

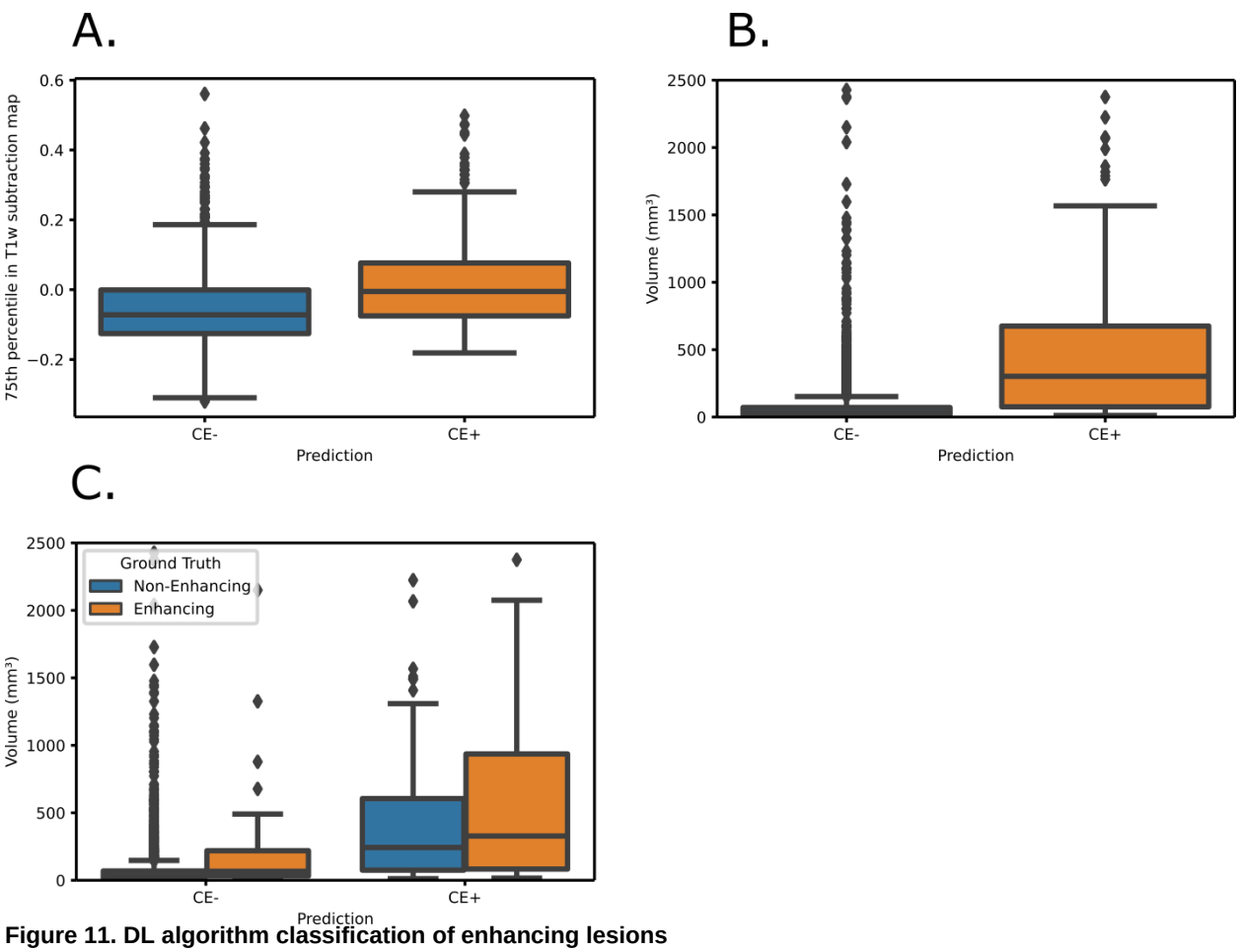


Figure 11. DL algorithm classification of enhancing lesions

A. In the test data set using a probability threshold of 0.75, lesions predicted to be enhancing by our DL algorithm had a significantly higher t1 subtraction signal. After selecting threshold, prediction data was binarized so that 0 represents non-enhancing lesions and 1 represents enhancing lesions. B. Analysis of additional lesion properties revealed that enhancing lesions in comparison to non-enhancing lesions are significantly larger in volume. C. Comparison with manually annotated images reveal that enhancing lesions that were predicted not be enhancing also have larger volumes.

Table 3. Confusion matrix on lesion level and patient level.

Sensitivity, Specificity and Accuracy of the predictions are shown below. CE-: Manually labelled as non-enhancing, CE+: Manually labelled as contrast-enhancing, Pred CE-: Predicted to be non-enhancing and Pred CE+: Predicted to be enhancing.

		Pred CE-	Pred CE+	Sensitivity	Specificity	Accuracy
Lesion-Level	**CE-**	3951	111	47.65%	97.27%	95.27%
	CE+	89	81			
Patient-level	**CE-**	86	44	82.76%	66.15%	71.28%
	CE+	10	48			

Validation with external data set from Lesjak et al., 2018

As a further validation of our prediction we used an external MSI dataset of 29 patients that was publicly available from the University of Ljubljana (Lesjak et al., 2018). ROC analysis of this dataset provided a similar AUC to our internal dataset with a value of 0.967 (Figure 12A). Further analysis of intensities and volumes of lesions from T1w subtraction maps revealed consistency with our internal test data (Figure 12B, 12C). Additionally, lesion-level sensitivity was calculated to be 93.33%, specificity to be 99.43% and accuracy to be 99.39%. On a patient level, sensitivity as determined to be 100%, specificity to be 75% and accuracy to be 82.76% (Table 4).

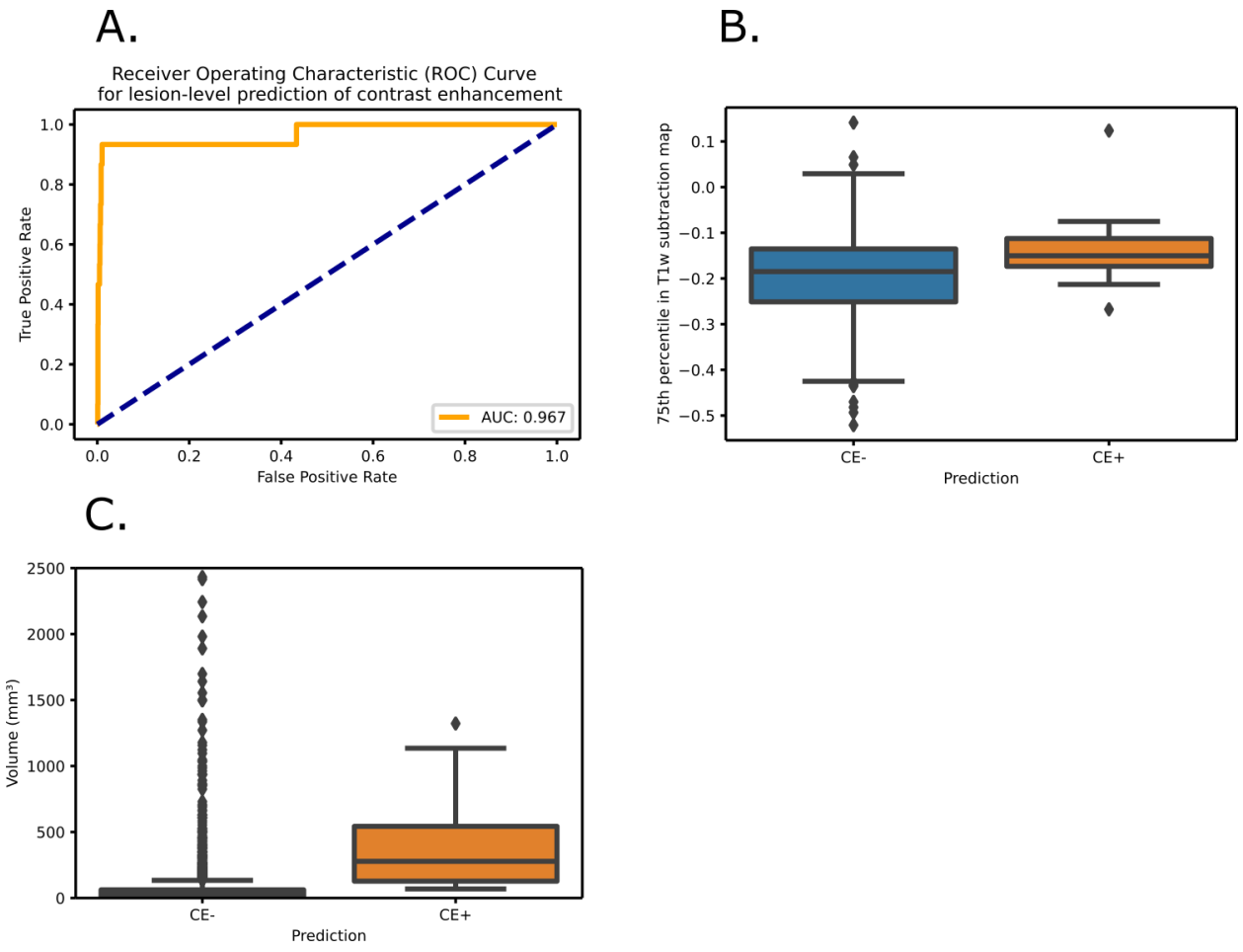


Figure 12. Validation of results using an external dataset *(Data from Lesjak et al., 2018).*

A. ROC analysis revealed an AUC of 0.967. Analysis of T1 Subtraction signals B, and lesion volumes, C, of predicted lesions reflected results of our internal test data.

Table 4. Confusion matrix using external dataset *(Data from Lesjak et al., 2018).*
Sensitivity, Specificity and Accuracy of the predictions are shown below: CE-: Manually labelled as non-enhancing, CE+: Manually labelled as contrast-enhancing, Pred CE-: Predicted to be non-enhancing and Pred CE+: Predicted to be enhancing.

		Pred CE-	Pred CE+	Sensitivity	Specificity	Accuracy
Lesion-Level	**CE-**	2276	13	93.33%	99.43%	99.39%
	CE+	1	14			
Patient-level	**CE-**	15	5	100%	75%	82.76%
	CE+	0	9			

In order to investigate the clinical value of prediction we conducted an exploratory analysis of lesion-wise contrast-enhancement from non-enhanced scans. For this purpose, we calculated multivariable logistic regression models to predict disease activity and progression (i.e. new or significantly enlarged lesion) in follow-up scans of patients. As exploratory variables, we used total number of lesions that were predicted to be contrast-enhancing (Table 4A) or the patient-wise absence or presence of lesions that were predicted to be enhancing (Table 4B). Considering the importance of lesion volume for the clinical disease severity and the prediction of our model, we used, in both models, total lesion volume as a covariate. A significant association was determined for both the lesion count predicted to be enhancing and the binary patient-wise label with imaging-assessed disease activity in the follow-up scan.

Table 5. Multivariable logistic regression model for predicting active disease and lesion count

Model A) Count of lesions predicted "enhancing"		
Variable	**Regression coefficient (95% CI)**	**p value**
Intercept	0.0058 (-0.381 - 0.393)	0.976
Total lesion volume (mm^3)	5.222e-5 (-7.21e-5 - 0.000)	0.410
Lesion count	0.3354 (0.004 - 0.667)	**0.048**

Model B) Presence of lesions predicted "enhancing" yes or no		
Variable	**Regression coefficient (95% CI)**	**p value**
Intercept	-0.0835 (-0.502 - 0.335)	0.696
Total lesion volume (mm^3)	7.909e-5 (-3.54e-5 - 0.000)	0.176
Enhancing lesions present	0.6734 (0.022 - 1.325)	**0.043**

Discussion

The MR assessment of disease progression plays a central role in the management of patients with Multiple Sclerosis, in particular the identification of contrast-enhancing (active) lesions. In this project we developed a deep learning architecture to predict the contrast-enhancement of lesions in MS patients from non-enhanced input FLAIR and T1w images. For this purpose a supervised deep learning strategy was employed. This approach required that enhancing lesions be manually annotated initially prior to running the DL algorithm. Through an assessment of the 75th percentile of these images, as proof of principle we showed, as expected, that contrast enhancing lesions display significantly higher T1w subtraction signals. Subsequently, these images were defined as ground truth for our DL network. For the purposes of training and testing the network, 3D patches around the center of mass of lesions were used. Plotting loss over epochs for quality-control showed that training was efficiently being performed. After training and validation, in a next step, we ran our network on the test data. An analysis of the accuracy of the test-run in correctly identifying enhancing lesions revealed a value of 95.27% on the lesion-level, and 71.28% on a patient-level. Furthermore, to investigate the clinical relevance of our findings, we showed that prediction of contrast-enhancing or active lesions indeed is significantly associated with disease activity in a patient's follow-up exam. In a final step to validate our results, we used an external dataset to test our network. The results of this analysis confirmed the findings from our internal dataset.

Although prediction of contrast enhancement has been performed previously at slice-level (Narayana et al., 2020), our study reveals the first study to perform this at a lesion-level. In contrast to the ability of slice-wise approaches to correctly predict disease activity on a patient level, our approach provides further insights in other lesion specific parameters such as the number and total volume of contrast-enhancing lesions. It is important to highlight that our approach avoids the requirement for a voxel-wise, exact segmentation of lesions, due to the patch-based cropping that occurs systematically around the center of mass of an individual lesion; This allows for even coarse annotations including single seed points to be sufficient for prediction. A manually segmentation of each lesion is then rendered unnecessary and instead only needs a detection of the lesion. Through our approach lesion features to be learned by the network were not pre-defined. This makes the approach unique in comparison to alternative radiomics based approaches, which require pre-defining of features as well as full lesion segmentation.

All lesions in MS typically follow a “life-cycle”, initiated by oedema and swelling and then subsequently followed by contrast enhancement a week later (Eisele et al., 2019; Eisele et al., 2012; Filippi et al., 1999). After approximately four weeks, there is a loss of contrast enhancement, followed by a decrease in swelling and oedema. Taking this temporal sequence of events into consideration and the close relationship between morphology and contrast enhancement several studies have attempted to use imaging methods to predict potential contrast-enhancing lesions from non-enhanced images (Gupta et al., 2017). Examples of these studies include fractional anisotropy where a parameter map is generated from diffusion tensor imaging and quantitative susceptibility mapping. Unlike these techniques which rely on advanced imaging techniques, our analysis utilized standard T1w and FLAIR images, which are part of a standard imaging protocol for MS (Traboulsee et al., 2016). The reliance on standard imaging protocols allows for a more widespread accessibility for implementation. In line with this argumentation, our model performed yielded similar results when applied to an external validation dataset when compared to our internal dataset. Although the cohort size of the external dataset was relatively smaller, the results suggest the applicability of our network to external data. It may however, be of interest to compare our method with the aforementioned advanced imaging techniques for the prediction of contrast enhancement, to investigate the added value for the prediction of contrast enhancement. As several factors in our approach may lead to an over- or underestimation of contrast enhancement, a combination with other approaches may improve clinical significance. Challenges in prediction may occur due to the variability in lifetime between lesions that are enhancing in comparison to oedema and lesion swelling. A shorter lifetime of contrast enhancement relative to oedema and lesion swelling would lead to an overestimation of the number of contrast-enhancing lesions when only relying solely on morphometric properties of lesions. It is also possible that oedema is undetectable in small lesions, which would then cause an underestimation of contrast enhancement in small lesions. Both of these occurrences were also found in our data with lesion size revealing to play an important role.

We decided to use a patch size of 16x16x16 mm (4096 mm^3), to optimize for enrichment of foreground information (i.e. lesion voxels over brain voxels) as well as to “fit” the majority of lesions. It was determined that enhancing lesions were larger in comparison to non-enhancing lesions, even when their mean volume was largely below that of the patch volume. The mean size of enhancing lesions was 476 mm^3. This caused the network to utilize lesion volume as a good surrogate marker for contrast-enhancement, and consequently this is one of the reasons that false-negative predictions were often seen with the smaller lesions. These findings validate

previous reports on the relationship of lesion load on clinical outcome (Filippi et al., 1999; Ge, 2006). In order to further optimize our network, a deeper exploration of complex augmentations such as heavy zooming and/or shearing or normalizing to an average lesion size, as opposed to the simple augmentations involving image geometry our current pipeline, could prove meaningful to learn additional properties of contrast-enhancing lesions.

To assess the potential of our network to predict enhancing lesions from non-enhanced scans for each patient in follow-up scans, we developed a multivariable logistic regression model. Upon testing of multiple variables, it was determined that the count of lesions as well as the presence of an enhancing lesions in the baseline scan was revealed to have a significant correlation with disease activity in the subsequent follow-up imaging. Interestingly, this correlation with disease activity was not present for both of these variables in the manually annotated dataset. This could indicate that the identified associations may have occurred as a result of the specific output generated by the deep learning algorithm, which would imply that improvements to the accuracy of the model need to be implemented before it is applied.

Deep Learning approaches used in indications and areas of medicine such as glioblastoma (Kleesiek et al., 2019) imaging and cardiovascular imaging (Bustamante et al., 2021), aimed to directly simulate contrast enhancing images have demonstrated that an unspecific synthesis of contrast-enhanced images is not possible (Lee et al., 2020). These results imply that a lesion-centered approach is necessary in the case of MS as presented here. Furthermore, DL techniques that prove to be successful in predicting contrast enhancement will lead to a significant reduction in contrast material needed for imaging MS patients (Gong et al., 2018).

There are several limitations in our study that affect the output of our network. As mentioned earlier, in our total dataset non-enhancing lesions outnumbered enhancing lesions approximately 10-fold. Though the use of oversampling techniques along with image augmentation allowed to partially mitigate this effect, a larger dataset with a greater number of enhancing lesions could improve the network output. Furthermore, the validation that we successfully performed using external data was conducted using a relatively small dataset. Thus, a further validation using larger and heterogenous external datasets could lead to better generalizability. Through our analysis, increased volume was found to be an important characteristic of enhancing lesions. This finding is substantiated by previous reports where it has been demonstrated that patients with high disease activity also reveal enlarging lesions followed

by shrinkage and/or growth over time (Köhler et al., 2019; Lee, 1998). We also found, interestingly, that lesion volume was associated with the number of false positives. Corrections to adjust for the effect of lesion volume would also improve the accuracy of the network. Similarly, adding more input sequences such as diffusion weighted imaging with improved lesion depiction would also improve the accuracy of classification. During the initial classification of lesions, we annotated their anatomic locations (periventricular, juxtacortical, infratentorial and subcortical). This anatomic information was, however, not subsequently used in our study as it was not within the scope of the project. There is substantial clinical value in understanding how enhancing lesions correlate with anatomic sites as this may improve our knowledge of Multiple Sclerosis and further our understanding of patients' symptoms.

Despite the wide contribution of ML approaches in transforming our understanding of phenomena in biology and medicine, the "Black Box" problem is an additional inherent limitation of the approach to be considered (Hayashi, 2019). Due to the way ML algorithms are designed, the only components that are transparent are the input variables and the output predictions. The remaining processes leading from input to output are not readily decipherable (Rudin & Radin, 2019). In a clinical context. this leads to physicians being provided with an output, but not a clear justification for the output (Price, 2018). Rendering DL algorithms more interpretable and modifiable would solve the black box problem and allow clinicians and users to troubleshoot as well as identify biases that may be present in the output (Rudin & Radin, 2019).

Ultimately, the strength of DL lies in combining radiological data with other clinical data of patients. Data including, but not restricted to patient symptoms, medication information, genomic data, clinical tests, pathological data and blood tests could strengthen the network leading to more accurate output. Figure 13 shows a schematic example of how such a workflow would look like where patient data (in this case for cancer patients) is integrated from different sources to determine a patient-specific diagnosis as well as personalized therapy plan (Bodalal et al., 2018). However, it has to be kept in mind that imaging is usually performed to objectify treatment decisions based on clinical findings. In such cases, an integration of data from other source may affect the network in a biased manner. Therefore, the appropriate validation must be in place to ensure that objectivity of the network output is maintained.

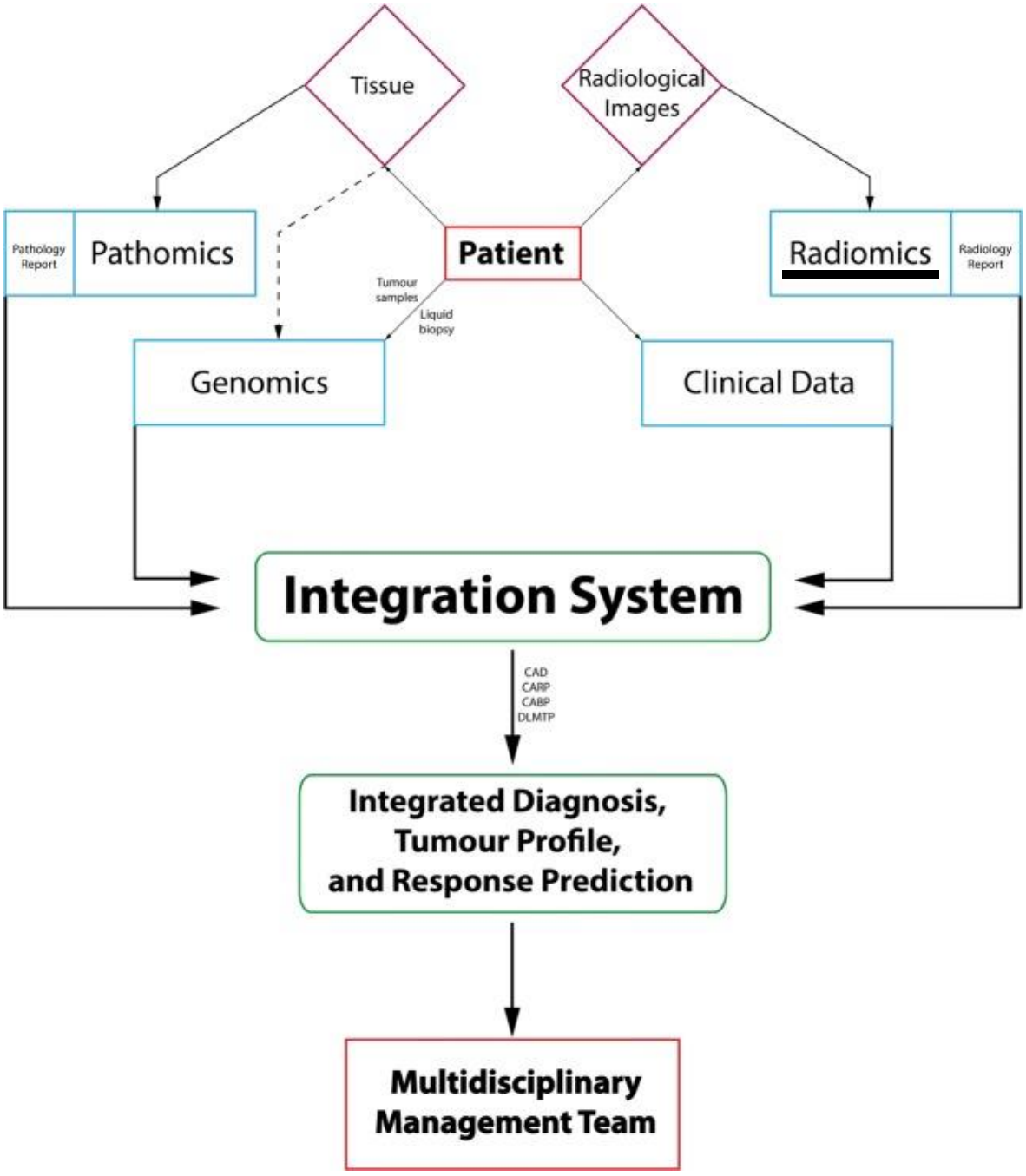


Figure 13. Model of interdisciplinary digital clinical workflow for personalized medicine.

The example above is for cancer patients, but the framework can be applied in principle to wide range of disease including Multiple Sclerosis. Image modified from *(Bodalal et al., 2018).*

Conclusions

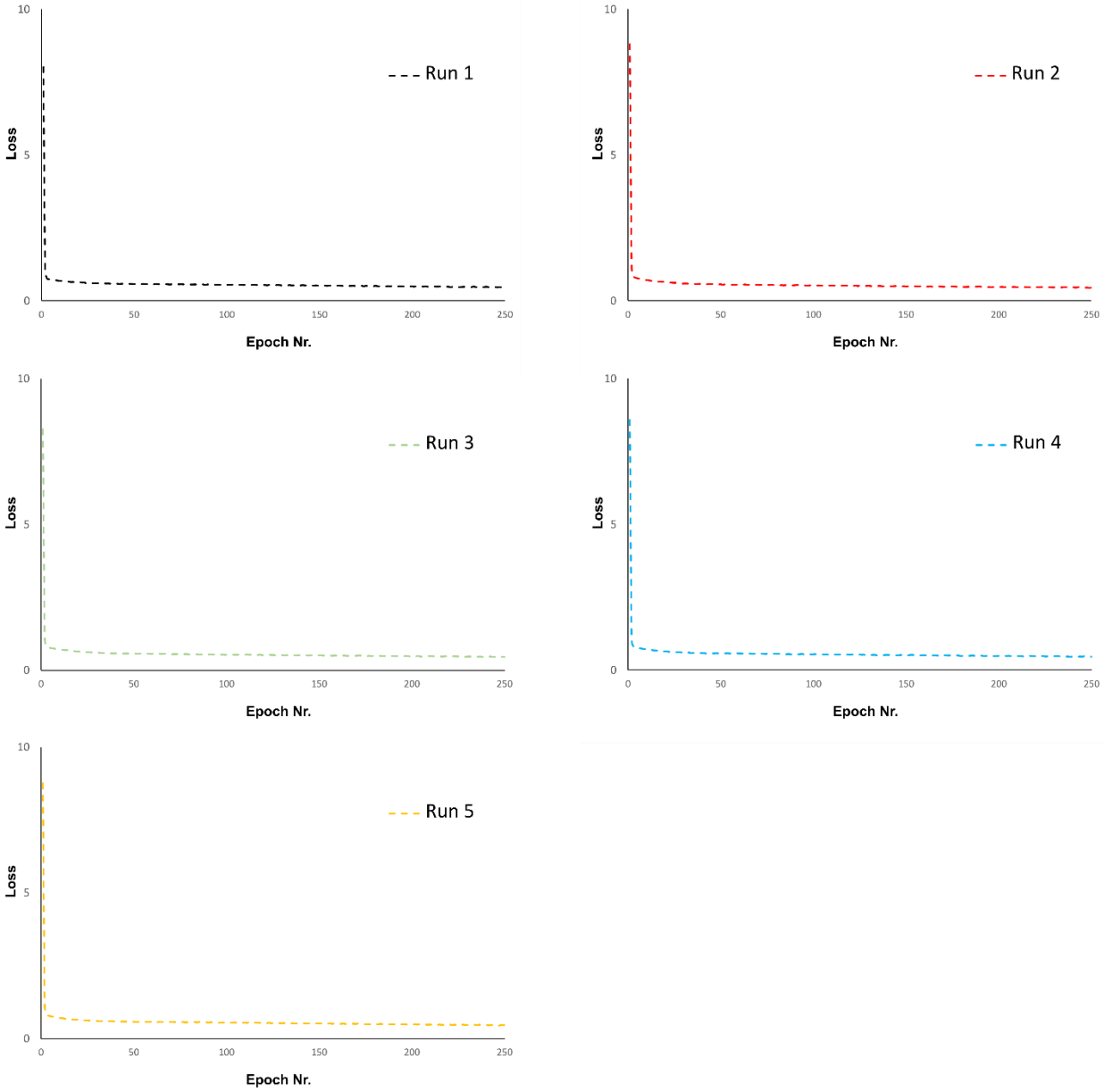
In summary, we developed a DL pipeline to predict lesions that are active and potentially contrast-enhancing without the need for administering GBCA. In the future, incorporating larger, more heterogeneous data sets could strengthen the accuracy of the network. Additional incorporation of non-imaging clinical data from patients could significantly improve the model's performance to not only predict lesion activity, but also clinical progression. Including pharmacological data and clinical symptoms, would allow the possibility to develop optimal treatment plans. For example, determining the relationship of specific medications on disease progress and relapse would further our understanding how best to optimally treat MS patients. Furthermore, incorporating genetic and family history data could help to identify genetic predispositions that are responsible for various forms of the disease. However, with each added layer of complexity to the network, the necessary validation steps must be conducted to avoid biased outputs from being generated. In the longterm, such an integrative model could serve in the future as an effective platform by playing an important role in long-term clinical management of MS and improving patient outcomes.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Benedikt Wiestler for his excellent supervision and guidance in the project as well as providing me with a strong introduction to the field of machine learning in radiology. His efforts helped immensely in bringing this project together and in organizing everything so that I could write my thesis. Additionally, I am very grateful to be a part of his team and for the wonderful discussions that we have had on the topic. I would like to thank all collaborators of this study, who are also co-authors in the manuscript that is based on the work of this thesis. These include: Timo Loehr, Matthias Bussas, Dominik Sepp, Lioba Grundl, Daria Bischl, Isabelle Riederer, Karolin Paprottka, Marie Metz, David Schinz, Christiane Gasperi, Achim Berthele, Sophia Grahl, Bernhard Hemmer, Claus Zimmer, Björn Menze, Mark Mühlau and Jan Kirschke. Special thanks to Dr. Jan Kirschke and Dr. Dominik Sepp for going through the basics of neuroradiographic image annotation. Further thanks to my mentor, Dr. Tobias Straub, who gave valuable advice and support through the thesis committee meetings. This work was supported by the German Research Foundation (DFG SPP2177, Radiomics: Next Generation of Biomedical Imaging) – project number 428223038. I would also like to thank DIFUTURE for their support of the study.

I am also deeply grateful to my parents, my brother and Saskia for their constant love and support. And of course, to my good Munich friends Alex, Mehdi and Connie a big thanks!

Appendix



Supplementary Figure 1. Loss curves of individual training runs.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems

.

Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26-41. https://doi.org/10.1016/j.media.2007.06.004

Barkhof, F., Scheltens, P., Frequin, S. T., Nauta, J. J., Tas, M. W., Valk, J., & Hommes, O. R. (1992). Relapsing-remitting Multiple Sclerosis: sequential enhanced MR imaging vs clinical findings in determining disease activity. *American Journal of Roentgenology*, *159*(5), 1041-1047. https://doi.org/10.2214/ajr.159.5.1414773

Bodalal, Z., Trebeschi, S., & Beets-Tan, R. (2018). Radiomics: a critical step towards integrated healthcare. *Insights into Imaging*, *9*(6), 911-914. https://doi.org/10.1007/s13244-018-0669-3

Brück, W., Bitsch, A., Kolenda, H., Brück, Y., Stiefel, M., & Lassmann, H. (1997). Inflammatory central nervous system demyelination: Correlation of magnetic resonance imaging findings with lesion pathology. *Annals of Neurology*, *42*(5), 783-793. https://doi.org/10.1002/ana.410420515

Bustamante, M., Viola, F., Carlhäll, C. J., & Ebbers, T. (2021). Using Deep Learning to Emulate the Use of an External Contrast Agent in Cardiovascular 4D Flow MRI. *Journal of Magnetic Resonance Imaging*. https://doi.org/10.1002/jmri.27578

Carpenter, K. A., Cohen, D. S., Jarrell, J. T., & Huang, X. (2018). Deep learning and virtual drug screening. *Future Medicinal Chemistry*, *10*(21), 2557-2567. https://doi.org/10.4155/fmc-2018-0314

Eichinger, P., Schön, S., Pongratz, V., Wiestler, H., Zhang, H., Bussas, M., Hoshi, M.-M., Kirschke, J., Berthele, A., Zimmer, C., Hemmer, B., Mühlau, M., & Wiestler, B. (2019). Accuracy of Unenhanced MRI in the Detection of New Brain Lesions in Multiple Sclerosis. *Radiology*, *291*(2), 429-435. https://doi.org/10.1148/radiol.2019181568

Eichinger, P., Wiestler, H., Zhang, H., Biberacher, V., Kirschke, J. S., Zimmer, C., Mühlau, M., & Wiestler, B. (2017). A novel imaging technique for better detecting new lesions in multiple sclerosis. *Journal of Neurology*, *264*(9), 1909-1918. https://doi.org/10.1007/s00415-017-8576-y

Eisele, P., Konstandin, S., Szabo, K., Ebert, A., Roßmanith, C., Paschke, N., Kerschensteiner, M., Platten, M., Schoenberg, S., Schad, L., & Gass, A. (2019). Temporal evolution of acute multiple sclerosis lesions on serial sodium (23Na) MRI. *Mult Scler Relat Disord.*, 48-54. https://doi.org/10.1016/j.msard.2019.01.027

Eisele, P., Szabo, K., Griebe, M., Roßmanith, C., Förster, A., Hennerici, M., & Gass, A. (2012). Reduced Diffusion in a Subset of Acute MS Lesions: A Serial Multiparametric MRI Study. *American Journal of Neuroradiology*, *33*(7), 1369-1373. https://doi.org/10.3174/ajnr.a2975

Errante, Y., Cirimele, V., Mallio, C. A., Di Lazzaro, V., Zobel, B. B., & Quattrocchi, C. C. (2014). Progressive increase of T1 signal intensity of the dentate nucleus on unenhanced magnetic resonance images is associated with cumulative doses of intravenously administered gadodiamide in patients with normal renal function, suggesting dechelation. *Invest Radiol.* , *49(10)*, 685-690. https://doi.org/10.1097/RLI.0000000000000072

Filippi, M., Rocca, M. A., Sormani, M. P., Pereira, C., & Comi, G. (1999). Short-term evolution of individual enhancing MS lesions studied with magnetization transfer imaging. *Magnetic Resonance Imaging*, *17*(7), 979-984. https://doi.org/10.1016/s0730-725x(99)00049-1

Ge, Y. (2006). Multiple Sclerosis: The Role of MR Imaging. *Am J Neuroradiol*, *27:1165–76*.

Gong, E., Pauly, J. M., Wintermark, M., & Zaharchuk, G. (2018). Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. *Journal of Magnetic Resonance Imaging*, *48*(2), 330-340. https://doi.org/10.1002/jmri.25970

Grahl, S., Pongratz, V., Schmidt, P., Engl, C., Bussas, M., Radetz, A., Gonzalez-Escamilla, G., Groppa, S., Zipp, F., Lukas, C., Kirschke, J., Zimmer, C., Hoshi, M., Berthele, A., Hemmer, B., & Mühlau, M. (2019). Evidence for a white matter lesion size threshold to support the diagnosis of relapsing remitting multiple sclerosis. *Mult Scler Relat Disord.*, 124-129. https://doi.org/10.1016/j.msard.2019.01.042

Gulani, V., Calamante, F., Shellock, F. G., Kanal, E., & Reeder, S. B. (2017). International Society for Magnetic Resonance in Medicine. Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol.*, *16(7)*, 564-570. https://doi.org/10.1016/S1474-4422(17)30158-8

Gupta, A., Al-Dasuqi, K., Xia, F., Askin, G., Zhao, Y., Delgado, D., & Wang, Y. (2017). The Use of Noncontrast Quantitative MRI to Detect Gadolinium-Enhancing Multiple Sclerosis Brain Lesions: A Systematic Review and Meta-Analysis. *American Journal of Neuroradiology*, *38*(7), 1317-1322. https://doi.org/10.3174/ajnr.a5209

Hayashi, Y. (2019). The Right Direction Needed to Develop White-Box Deep Learning in Radiology, Pathology, and Ophthalmology: A Short Review. *Frontiers in Robotics and AI*, *6*. https://doi.org/10.3389/frobt.2019.00024

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017, 2017-07-01). Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),

Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H. P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K. H., & Kickingereder, P. (2019). Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping*, *40*(17), 4952-4964. https://doi.org/10.1002/hbm.24750

Jonkman, L. E., Soriano, A. L., Amor, S., Barkhof, F., Van Der Valk, P., Vrenken, H., & Geurts, J. J. G. (2015). Can MS lesion stages be distinguished with MRI? A postmortem MRI and histopathology study. *Journal of Neurology*, *262*(4), 1074-1080. https://doi.org/10.1007/s00415-015-7689-4

Kanal, E., Maravilla, K., & Rowley, H. A. (2014). Gadolinium Contrast Agents for CNS Imaging: Current Concepts and Clinical Evidence. *American Journal of Neuroradiology*, *35*(12), 2215-2226. https://doi.org/10.3174/ajnr.a3917

Kanda, T., Ishii, K., Kawaguchi, H., Kitajima, K., & Takenaka, D. (2014). High Signal Intensity in the Dentate Nucleus and Globus Pallidus on Unenhanced T1-weighted MR Images: Relationship with Increasing Cumulative Dose of a Gadolinium-based Contrast Material. *Radiology*, *270*(3), 834-841. https://doi.org/10.1148/radiol.13131669

Kanda, T., Oba, H., Toyoda, K., Kitajima, K., & Furui, S. (2016). Brain gadolinium deposition after administration of gadolinium-based contrast agents. *Japanese Journal of Radiology*, *34*(1), 3-9. https://doi.org/10.1007/s11604-015-0503-5

Kleesiek, J., Morshuis, J. N., Isensee, F., Deike-Hofmann, K., Paech, D., Kickingereder, P., Köthe, U., Rother, C., Forsting, M., Wick, W., Bendszus, M., Schlemmer, H. P., & Radbruch, A. (2019). Can Virtual Contrast Enhancement in Brain MRI Replace Gadolinium?: A Feasibility Study. *Invest Radiol.* (Oct;54(10)), 653-660.

Köhler, C., Wahl, H., Ziemssen, T., Linn, J., & Kitzler, H. H. (2019). Exploring individual multiple sclerosis lesion volume change over time: Development of an algorithm for the analyses of longitudinal quantitative MRI measures. *NeuroImage: Clinical*, *21*, 101623. https://doi.org/10.1016/j.nicl.2018.101623

Lee, D., Moon, W. J., & Ye, J. C. (2020). Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks. *Nat Mach Intell*, *2*, 34–42

Lee, M. (1998). Defining multiple sclerosis disease activity using MRI T2-weighted difference imaging. *Brain*, *121*(11), 2095-2102. https://doi.org/10.1093/brain/121.11.2095

Lesjak, Ž., Galimzianova, A., Koren, A., Lukin, M., Pernuš, F., Likar, B., & Špiclin, Ž. (2018). A Novel Public MR Image Dataset of Multiple Sclerosis Patients With Lesion Segmentations Based on Multi-rater Consensus. *Neuroinformatics*, *16*(1), 51-63. https://doi.org/10.1007/s12021-017-9348-7

Li, D., Held, U., Petkau, J., Daumer, M., Barkhof, F., Fazekas, F., Frank, J., Kappos, L., Miller, D., Simon, J., Wolinsky, J., Filippi, M., & Lawry, S. (2006). MRI T2 lesion burden in multiple sclerosis: a plateauing relationship with clinical disability. *Neurology*, *May 9;66(9):1384-9*. https://doi.org/10.1212/01.wnl.0000210506.00078.5c

Li, H., Jiang, G., Zhang, J., Wang, R., Wang, Z., Zheng, W.-S., & Menze, B. (2018). Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images. *NeuroImage*, *183*, 650-665. https://doi.org/10.1016/j.neuroimage.2018.07.005

Lublin, F. D., & Reingold, S. C. (1996). Defining the clinical course of multiple sclerosis: Results of an international survey. *Neurology*, *46*(4), 907-911. https://doi.org/10.1212/wnl.46.4.907

Mcdonald, R. J., Mcdonald, J. S., Kallmes, D. F., Jentoft, M. E., Murray, D. L., Thielen, K. R., Williamson, E. E., & Eckel, L. J. (2015). Intracranial Gadolinium Deposition after Contrast-enhanced MR Imaging. *Radiology*, *275*(3), 772-782. https://doi.org/10.1148/radiol.15150025

Mcdonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F. D., Mcfarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C., Sandberg-Wollheim, M., Sibley, W., Thompson, A., Van Den Noort, S., Weinshenker, B. Y., & Wolinsky, J. S. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology*, *50*(1), 121-127. https://doi.org/10.1002/ana.1032

McGinley, M. P., Goldschmidt, C. H., & Rae-Grant, A. D. (2021). Diagnosis and Treatment of Multiple Sclerosis. *JAMA*(325(8)), 765–779. https://doi.org/10.1001/jama.2020.26858

Narayana, P. A., Coronado, I., Sujit, S. J., Wolinsky, J. S., Lublin, F. D., & Gabr, R. E. (2020). Deep Learning for Predicting Enhancing Lesions in Multiple Sclerosis from Noncontrast MRI. *Radiology*, *294*(2), 398-404. https://doi.org/10.1148/radiol.2019191061

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., Mclean, C. Y., & Depristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, *36*(10), 983-987. https://doi.org/10.1038/nbt.4235

Price, W. N. (2018). Big data and black-box medical algorithms. *Science Translational Medicine*, *10*(471), eaao5333. https://doi.org/10.1126/scitranslmed.aao5333

Reich, D. S., Lucchinetti, C. F., & Calabresi, P. A. (2018). Multiple Sclerosis. *New England Journal of Medicine*, *378*(2), 169-180. https://doi.org/10.1056/nejmra1401483

Rogosnitzky, M., & Branch, S. (2016). Gadolinium-based contrast agent toxicity: a review of known and proposed mechanisms. *BioMetals*, *29*(3), 365-376. https://doi.org/10.1007/s10534-016-9931-7

Rudin, C., & Radin, J. (2019). Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition. *1.2*, *1*(2). https://doi.org/10.1162/99608f92.5a8a3a3d

Saade, C., Bou-Fakhredin, R., Yousem, D. M., Asmar, K., Naffaa, L., & El-Merhi, F. (2018). Gadolinium and Multiple Sclerosis: Vessels, Barriers of the Brain, and Glymphatics. *American Journal of Neuroradiology*, *39*(12), 2168-2176. https://doi.org/10.3174/ajnr.a5773

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V. J., Zimmer, C., Hemmer, B., & Mühlau, M. (2012). Gadolinium deposition in the brain: summary of evidence and recommendations. *NeuroImage*(59(4)), 3774-3783. https://doi.org/10.1016/j.neuroimage.2011.11.032.

Schmidt, P., Pongratz, V., Küster, P., Meier, D., Wuerfel, J., Lukas, C., Bellenberg, B., Zipp, F., Groppa, S., Sämann, P. G., Weber, F., Gaser, C., Franke, T., Bussas, M., Kirschke, J., Zimmer, C., Hemmer, B., & Mühlau, M. (2019). Automated segmentation of changes in FLAIR-hyperintense white matter lesions in multiple sclerosis on serial magnetic resonance imaging. *NeuroImage: Clinical*, *23*, 101849. https://doi.org/10.1016/j.nicl.2019.101849

Stephen, L., Hauser, S., & Josephson, A. (2016). *Harrison's Neurology in Clinical Medicine* (t. Edition, Ed.). McGraw-Hill Education.

Swaminathan, S., High, W. A., Ranville, J., Horn, T. D., Hiatt, K., Thomas, M., Brown, H. H., & Shah, S. V. (2008). Cardiac and vascular metal deposition with high mortality in nephrogenic systemic fibrosis. *Kidney International*, *73*(12), 1413-1418. https://doi.org/10.1038/ki.2008.76

Tafti, D., Ehsan, M., & Xixis, K. (2021). *Multiple Sclerosis*. Statpearls.

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., Mowry, E. M., Sorensen, P. S., Tintoré, M., Traboulsee, A. L., Trojano, M., Uitdehaag, B. M. J., Vukusic, S., Waubant, E., Weinshenker, B. G., Reingold, S. C., & Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, *17*(2), 162-173. https://doi.org/10.1016/s1474-4422(17)30470-2

Traboulsee, A., Simon, J. H., Stone, L., Fisher, E., Jones, D. E., Malhotra, A., Newsome, S. D., Oh, J., Reich, D. S., Richert, N., Rammohan, K., Khan, O., Radue, E.-W., Ford, C., Halper, J., & Li, D. (2016). Revised Recommendations of the Consortium of MS Centers Task Force for a Standardized MRI Protocol and Clinical Guidelines for the Diagnosis and Follow-Up of Multiple Sclerosis. *American Journal of Neuroradiology*, *37*(3), 394-401. https://doi.org/10.3174/ajnr.a4539

Tullman, M. J. (2013). Overview of the epidemiology, diagnosis, and disease progression associated with multiple sclerosis. *Am J Manag Care*, S15-20.

Van Der Valk, P., & De Groot, C. J. A. (2000). Staging of multiple sclerosis (MS) lesions: pathology of the time frame of MS. *Neuropathology and Applied Neurobiology*, *26*(1), 2-10. https://doi.org/10.1046/j.1365-2990.2000.00217.x

Webb, S. (2018). Deep learning for biology. *Nature*, *554*(7693), 555-557. https://doi.org/10.1038/d41586-018-02174-z

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, *31*(3), 1116-1128. https://doi.org/10.1016/j.neuroimage.2006.01.015

Publications

Current Thesis Project:

Sasidharan N., Löhr T., Bussas M., Sepp D., Grundl L., Bischl D., Riederer I., Paprottka K., Metz M., Schinz D., Gasperi C., Berthele A., Grahl S., Hemmer B., Zimmer C., Menze B., Mühlau M., Kirschke J. and Wiestler B., (2023) A lesion-level deep learning approach to predict enhancing lesions from non-enhanced images in Multiple Sclerosis" **Manuscript in preparation.**

Other Projects:

Gong, Y., **Sasidharan, N.,** Laheji, F., Frosch, M., Musolino, P., Tanzi, R., Kim, D. Y., Biffi, A., El Khoury, J. and Eichler, F. (2017), Microglial dysfunction as a key pathological change in adrenomyeloneuropathy. Ann Neurol., 82: 813–827.

Pym, E., **Sasidharan, N.,** Thompson-Peer, K.L., Simon, D.J., Anselmo, A., Sadreyev, R., Hall, Q., Nurrish, S., and Kaplan, J.M. (2017). Shank is a dose-dependent regulator of Cav1 calcium current and CREB target expression. eLife 6.

Ailion M., Hannemann M., Dalton S., Pappas A., Watanabe S., Hegermann J., Liu Q., Han HF., Gu M,. Goulding M., **Sasidharan N**., et al. (2014) "Two Rab2 Interactors Regulate Dense-Core Vesicle Maturation". Neuron 82 167-180.

Sasidharan, N., Sumakovic, M., Hannemann, M., Hegermann, J., Liewald, J.F., Olendrowitz, C., Koenig, S., Grant, B.D., Rizzoli, S.O., Gottschalk, A., and Eimer, S. (2012). "RAB-5 and RAB-10 cooperate to regulate neuropeptide release in Caenorhabditis elegans. " Proceedings of the National Academy of Sciences of the United States of America *109*, 18944-18949.

Hannemann, M., **Sasidharan, N**., Hegermann, J., Kutscher, L.M., Koenig, S., and Eimer, S. (2012). "TBC-8, a putative RAB-2 GAP, regulates dense core vesicle maturation in Caenorhabditis elegans." PLoS genetics 8, e1002722.

Luo, L., Hannemann, M., Koenig, S., Hegermann, J., Ailion, M., Cho, M.K., **Sasidharan, N.,** Zweckstetter, M., Rensing, S.A., and Eimer, S. (2011). "The Caenorhabditis elegans GARP complex contains the conserved Vps51 subunit and is required to maintain lysosomal morphology." Molecular biology of the cell 22, 2564-2578.